D

AMERICAN
JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXXIII, Number 3
JULY, 1951

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

CONTENTS

The American Journal of Mathematics will appear four times yearly.

The subscription price of the Journal for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the Journal remain on sale.

Papers intended for publication in the Journal may be sent to any of the Editors.

Editorial communications should be sent to Professor Aurel Wintner at The Johns Hopkins University.

Subscriptions to the Journal and all business communications should be sent to The Johns Hopkins Press, Baltimore 18, Maryland, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

A THEORY OF DEGREE OF MAPPING BASED ON INFINITESIMAL ANALYSIS.*

By Mitio Nagumo.

1. Introduction. This paper establishes a theory of degree of mapping for open sets in a Euclidean space of finite dimension, based on the theory of infinitesimal analysis, which is free from the notion of simplicial mapping. Although the results are not new, I hope in this way to make it possible to incorporate the theory of degree of mapping into a course in infinitesimal analysis.

A mapping $x' = f(x) = \{f_i(x) \mid i = 1, \cdots, m\}$ $(x = (x_1, \cdots, x_m))$ of a bounded open set G of an m-dimensional Euclidean space E^m into E^m is called *regular* on G, if each $f_i(x)$ is continuously differentiable on the closure $\bar{G}$ of G. We write $\partial x_i f$ or $\partial_i f$ for $\partial f/\partial x_i$, and by $D(f_1, \cdots, f_m/x_1, \cdots, x_m)$, or more concisely by $D(f/x)$, we denote the functional determinant $\det(\partial_i f_j)_{i,j=1,\ldots,m}$. We call the point x, for which $D(f/x)$ vanishes, a *critical point* of f; and the image of the set of all critical points (of f) by f will be called the *crease* of f on G. The expression $a \bar{\varepsilon} A$ means "a does not belong to A."

Now let a be a point of E^m which lies neither on $f(\bar{G} - G)$ (the image of the boundary of G) nor on the crease of f. Then, by a theorem on implicit functions, there exist at most a finite number of points x in G such that $f(x) = a$. Let the number of those points for which $D(f/x) > 0$ be p, and the number of those for which $D(f/x) < 0$ be q. Then we call the integer $p - q$ the *degrees of mapping* of G at a by f, and denote it by

$$A[a, G, f] \,(= p - q).$$

We have the following properties of the degree of mapping:

i) *If f is the identical mapping $f(x) = x$, then*

$A[a, G, f] = 1$ *when* $a \varepsilon G$, $A[a, G, f] = 0$ *when* $a \bar{\varepsilon} \bar{G}$.

ii) *If $A[a, G, f] \neq 0$ and has a meaning, then there exists a point $x \varepsilon G$, such that $f(x) = a$.*

* Received March 6, 1950; revised September 4, 1950.

Essentially the same treatise by the author was published in a brochure in the Japanese language with examples and applications, in which (following Leray and Schauder) the theory is extended to a class of mappings in Banach space.

iii) *If G is divided into open sets G_i $(i=1,\cdots,k)$ without common points, i. e., $G \supset G_1 \cup G_2 \cup \cdots \cup G_k$, $\bar{G} = \bar{G}_1 \cup \bar{G}_2 \cup \cdots \cup \bar{G}_k$ and $G_i \cap G_j = 0$ (empty set) $(i \neq j)$, then*

$$A[a, G, f] = \sum_{i=1}^{k} A[a, G_i, f],$$

if every term has a meaning.

iv) *If $f_t(x)$ is a continuous function of (t, x) for $0 \leqq t \leqq 1$, $x \in \bar{G}$, and $a(t) (\varepsilon E^m)$ is continuous and $a(t) \bar{\varepsilon} f_t(\bar{G} - G)$ for all t in $0 \leqq t \leqq 1$, then $A[a(t), G, f_t]$ is constant for $0 \leqq t \leqq 1$.*

The propositions i), ii), and iii) are obvious; but for the proof of iv) we need considerable preparation. In § 2 we shall give auxiliary theorems for the proof of iv) which will be carried out in § 3. In § 4 the definition of degree of mapping will be extended to any continuous mapping of an open set G in E^m. In § 5 a product theorem of degree of mapping, concerning the composition of two mappings, will be proved.

2. Auxiliary theorems.

THEOREM 1.[1] *Let $x' = f(x)$ be a regular mapping of an open set G in E^m. Then the crease of f is a set of (m-dimensional Lebesgue) measure zero in E^m.*

Proof. We divide G into an enumerable set of closed cubes Q: $|x_i - a_i| \leqq l$ $(i = 1, 2, \cdots, m)$. Then it is sufficient to prove that the measure of the crease of f on Q is zero. For any $\epsilon > 0$, there exists a natural number n such that Q can be divided into n^m equal small closed cubes Q_ν $(\nu = 1, \cdots, n^m)$ in each of which we have

$$f_i(x) = f_i(\alpha) + \sum_{j=1}^{m} \partial_j f_i(\alpha)(x_j - \alpha_j) + \eta_i(x),$$

where $\sum_{i=1}^{m} \eta^2 < (\epsilon/n)^2$ and α is an arbitrary point in Q_ν. Let $\hat{Q}_\nu$ be the image of Q_ν under the linear transformation $x'_i = f_i(\alpha) + \sum_{j=1}^{m} \partial_j f_i(\alpha)(x_j - \alpha_j)$. The diameter of $\hat{Q}_\nu$ is not greater than $2Ll/n$, where $L = \mathrm{Max}_Q(\sum_{i,j=1}^{m} \{\partial_j f_i(x)\}^2)^{1/2}$. If Q_ν contains a critical point, we take it as α. Then $\hat{Q}_\nu$ is contained in an

[1] This is the simplest special case of a theorem of A. Sard, *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 883-890.

$(m-1)$-dimensional linear manifold, and $f(Q_\nu)$ lies in the ϵ/n-neighborhood of $\hat{Q}_\nu$. Therefore the measure of $f(Q_\nu)$ is smaller than $\{2n^{-1}(Ll+\epsilon)\}^{m-1}2\epsilon n^{-1}$. Consequently the measure of the crease in $f(Q)$ is smaller than $2^m(Ll+\epsilon)^{m-1}\epsilon$, and since ϵ is arbitrary small, it must be zero.

As an application of Theorem 1 we obtain immediately:

THEOREM 2. *Let M be any set of points in E^m which can be represented by the equations*

$$x_\mu = \phi_\mu(s_1,\cdots,s_l) \qquad (\mu=1,\cdots,m), \quad 0<l<m,$$

where $\phi_\mu(s_1,\cdots,s_l)$ are continuously differentiable in an open domain of an $l(<m)$-dimensional Euclidean space. Then M is a set of (m-dimensional Lebesgue) measure zero in E^m.

THEOREM 3. *Let G be a bounded open set in E^m and $x'=F(x)$ be a regular mapping of G into E^m. Then for any point $a \in E^m$, neither on $F(\bar{G}-G)$ nor on the crease of F, there exists a positive number ϵ with the following property:*

For any regular mapping $x'=f(x)$ of G into E^m such that $|f(x)-F(x)|<\epsilon^2$ for $x\in\bar{G}$ and $|\partial_j f(x)-\partial_j F(x)|<\epsilon$ $(j=1,\cdots,m)$ for $x\in G_\epsilon$, where $G_\epsilon = G\cap\{x \mid \mathrm{dist}(x,\bar{G}-G)>\epsilon\}$, the equality $A[a,G,f] = A[a,G,F]$ holds.

Proof. There are only a finite number of points x of G such that $F(x)=a$. Let $p^1,\cdots,p^N$ be all such points. Let K be the set of critical points of F, and put

$$\Delta = \mathrm{Min}\{|p^\mu - p^\nu| \; (\mu\neq\nu), \quad \mathrm{dist}(p^\mu, K\cup(\bar{G}-G))\};$$

then $\Delta>0$. Let α be a positive number such that $\alpha<\Delta/2$, and $U_\alpha(p^\nu)$ be the α-neighborhood of p^ν, and put

$$M_\alpha = \mathrm{Min}\{|F(x)-a| \mid x\in\bar{G}-\bigcup_{\nu=1}^{N}U_\alpha(p^\nu)\};$$

then $U_\alpha(p^\nu)\subset G_\alpha$, where $G_\alpha = G\cap\{x\mid \mathrm{dist}(x,\bar{G}-G)>\alpha\}$, and $M_\alpha>0$.

Since $\det(\partial_x F(x))\neq 0$ for $x\in\bar{U}_\alpha(p^\nu)$, there exists an $\epsilon'>0$ such that $\det(\partial_x f(x))\neq 0$ holds for any regular mapping f of $U_\alpha(p^\nu)$ into E^m provided that $|\partial_j f(x)-\partial_j F(x)|<\epsilon'$ $(j=1,\cdots,m)$.

Now let f be any regular mapping of G into E^m such that $|f(x)-F(x)|$

---

[2] By $|p-q|$ we denote the distance $(\sum_{i=1}^{m}(p_i-q_i)^2)^{1/2}$ of two points p and q in E^m.

$< M_\alpha$ for $x \varepsilon \bar{G}$ and $|\partial_j f(x) - \partial_j F(x)| < \epsilon'$ $(j = 1, \cdots, m)$ for $x \varepsilon G_\alpha$. We put $f_t(x) = (1 - t)F(x) + tf(x)$. Then we have

$$\det(\partial_x f_t(x)) \neq 0 \text{ for } 0 \leqq t \leqq 1, x \varepsilon U_\alpha(p^\nu),$$

and all roots of $f_t(x) = a$ lie in $\bigcup_{\nu=1}^{N} U_\alpha(p^\nu)$.

Let τ be any value of t from $0 \leqq t \leqq 1$ and p_0 be a point of G such that $f_\tau(p_0) = a$. Then p_0 belongs to a $U_\alpha(p^\nu)$ and by the theory of implicit functions there exists a neighborhood of $t = \tau$ such that the equation $f_t(x) = a$ has exactly one solution $x = p(t)$ in a sufficiently small neighborhood of p_0. This solution $p(t)$ depends on t continuously and can be prolongated for the whole interval $0 \leqq t \leqq 1$. In fact, let $p(t)$ be continuous and $p(t) \varepsilon U_\alpha(p^\nu)$ for $\tau \leqq t < T$, and p^* be any accumulation point of $p(t)$ for $t \to T$. Then $p^* \varepsilon U_\alpha(p^\nu)$ and $f_T(p^*) = a$. Since $\det(\partial_x f_T(p^*)) \neq 0$, then by the theory of implicit functions there exists a unique solution $x = p^*(t)$ of $f_t(x) = a$ in a sufficiently small neighborhood of p^* for t sufficiently near to T. As $p^*(t)$ is continuous and the solution of $f_t(x) = a$ is unique in a small neighborhood of p^*, it must coincide with $p(t)$ for $t < T$ sufficiently near to T. Thus $p(t)$ is continuous for $\tau \leqq t \leqq T$ and can be prolongated beyond T if $T < 1$. This and a similar consideration for $t \leqq \tau$ show that $p(t)$ can be prolongated for $0 \leqq t \leqq 1$, where it is continuous. Since the solutions of $f_t(x) = a$ are isolated and continuous for $0 \leqq t \leqq 1$, the number of the solutions of $f_t(x) = a$ is constant in each $U_\alpha(p^\nu)$ for $0 \leqq t \leqq 1$. Hence $f_t(x) = a$ has exactly one solution in each $U_\alpha(p^\nu)$ and $\det(\partial_x f_t)$ keeps the same sign in each $U_\alpha(p^\nu)$ for $0 \leqq t \leqq 1$. Then $\Delta[a, G, f_t]$ is constant for $0 \leqq t \leqq 1$. The proof is thus established, if we take $\epsilon = \mathrm{Min}\{\alpha, M_\alpha, \epsilon'\}$.

THEOREM 4. *Let G be a bounded open set in E^m, $x' = F(x)$ a regular mapping of G into E^m, $a(t) \varepsilon E^m$ continuously differentiable in $0 \leqq t \leqq 1$, and $a(t) \bar{\varepsilon} F(\bar{G} - G)$. Then there exists for any $\epsilon > 0$ such that $\epsilon < \mathrm{dist}(a(t), F(\bar{G} - G))$ a regular mapping $x' = f(x)$ of G into E^m with the following properties:*

i) $|f(x) - F(x)| < \epsilon$ *for* $x \varepsilon \bar{G}$ *and* $|\partial_j f(x) - \partial_j F(x)| < \epsilon$ $(j = 1, \cdots, m)$ *for* $x \varepsilon G_\epsilon$.[3]

ii) *For every x such that $f(x) = a(t)$ $(0 \leqq t \leqq 1)$, there holds the inequality*

$$\text{Rank of } (\partial_x f(x)) \geqq m - 1, \textit{ where } (\partial_x f) = (\partial_j f_i)_{i,j=1,\ldots,m}.$$

[3] $G_\epsilon = G \cap \{x \mid \mathrm{dist}(x, \bar{G} - G) > \epsilon\}$.

iii) *For any fixed t the set of points x such that $f(x) = a(t)$ has no accumulation point in G.*

Proof. We can find a regular mapping $x' = \phi(x)$ of G into E^m, such that $\phi(x)$ is 3 times continuously differentiable in G, $|\phi(x) - F(x)| < \epsilon/2$ for $x \varepsilon \bar{G}$ and $|\partial_j\phi(x) - \partial_j F(x)| < \epsilon/2$ for $x \varepsilon G_\epsilon$. We put

$$(1) \qquad f_i(x) = \phi_i(x) + \alpha_i + \sum_{\nu=1}^{m} (\beta_{i,\nu}x_\nu + 2^{-1}\gamma_{i,\nu}x^2{}_\nu).$$

Then we can find $\delta > 0$, such that $f(x)$ defined by (1) satisfies the statement i) if the conditions $|\alpha_i| < \delta$, $|\beta_{i,\nu}| < \delta$, $|\gamma_{i,\nu}| < \delta$ are satisfied.

Let us prove that the set of (α, β, γ) such that $f(x)$ defined by (1) does not satisfy ii) has the measure zero in $E^q{}_{(\alpha,\beta,\gamma)}$ $(q = m + 2m^2)$. If ii) does not hold, then there exists a point x and a value of t for which $f(x) = a(t)$ and certain $l \geqq 2$ columns of the matrix $(\partial_x f)$ are linear combinations of the other $m - l$ columns. Let us suppose that these are the last l columns. Then there exist $l(m-l)$ numbers $\lambda_{\mu,\nu}$ $(\mu = 1, \cdots, l,$ $\nu = 1, \cdots, m - l)$ such that

$$\partial_{m-l+\mu} f_i = \sum_{\nu=1}^{m-l} \lambda_{\mu,\nu}\partial_\nu f_i \qquad (i = 1, \cdots, m).$$

Then by (1):

$$\beta_{i,m-l+\mu} = \sum_{\nu=1}^{m-l} \lambda_{\mu,\nu}(\beta_{i,\nu} + \partial_\nu\phi_i(x) + \gamma_{i,\nu}x_\nu)$$
$$- \partial_{m-l+\mu}\phi_i(x) - \gamma_{i,m-l+\mu}x_{m-l+\mu},$$

and

$$\alpha_i = a_i(t) - \phi_i(x) - \sum_{\nu=1}^{m} (\beta_{i,\nu}x_\nu + 2^{-1}\gamma_{i,\nu}x_\nu{}^2).$$

This shows that (α, β, γ) can be represented by continuously differentiable functions of $m + 2m^2 - l^2 + 1 < q$ arguments $t, x_j, \lambda_{\mu,\nu}, \beta_{i,\nu}, \gamma_{i,\nu}$ $(i, j = 1,$ $\cdots, m, \mu = 1, \cdots, l, \nu = 1, \cdots, m - l)$. Thus by Theorem 2 the set of points (α, β, γ) such that $f(x)$ does not satisfy ii) has the measure zero in $E^q{}_{(\alpha,\beta,\gamma)}$, since there are only a finite number of ways to choose $l \geqq 2$ columns from $(\partial_x f)$.

Next we shall prove that the set of (α, β, γ) for which the statement ii) holds, but not iii), has also the measure zero in $E^q{}_{(\alpha,\beta,\gamma)}$. Suppose that this occurs for $t = t^*$ and put $a(t^*) = a$. Then the equation $f(x) = a$ has an infinite number of roots $x = p^n = (p_1{}^n, \cdots, p_m{}^n)$, which converges to a point $p = (p_1, \cdots, p_m) \varepsilon G$. By ii) there exist $m - 1$ components of $f(x)$, say $f_1, \cdots, f_{k-1}, f_{k+1}, \cdots, f_m$, such that $D(f_1, \cdots, f_{k-1}, f_{k+1}, \cdots, f_m/x_1,$

$\cdots, x_{m-1}) \neq 0$ for $x = p$. If we put $y_i = f_i(x)$ for $i \neq k$ and $y_k = x_m$, then $D(y/x)_{x=p} = \pm D(f_1, \cdots, f_{k-1}, f_{k+1}, \cdots, f_m/x_1, \cdots, x_{m-1}) \neq 0$, and by a theorem of implicit functions, a neighborhood of p in $E_{(x)}{}^m$ corresponds homeomorphically to a neighborhood of $(a_1, \cdots, a_{k-1}, p_m, a_{k+1}, \cdots, a_m)$ in $E_{(y)}{}^m$. Let $x = \psi(y)$ be the inverse mapping (3 times continuously differentiable) and put $p_m{}^n = s_n$, $p_m = \sigma$, $\psi_\mu(a_1, \cdots, a_{k-1}, s, a_{k+1}, \cdots, a_m) = h_\mu(s)$ $(\mu = 1, \cdots, m)$. Then we have $\lim s_n = \sigma$, $p_\mu{}^n = h_\mu(s_n)$, $p_\mu = h_\mu(\sigma)$ and $f_i(h(s_n)) = f_i(h(\sigma)) = a_i$. Thus, if we put $h'_\mu(\sigma) = \lambda_\mu$ and $h''_\mu(\sigma) = \lambda'_\mu$, we have

$$(2) \qquad \frac{d}{ds} f_i(h(s))_{s=\sigma} = \sum_{\nu=1}^{m} \lambda_\nu (\partial_\nu f_i)_{x=p} = 0,$$

and

$$(3) \qquad \frac{d^2}{ds^2} f_i(h(s))_{s=\sigma} = \sum_{\mu,\nu=1}^{m} \lambda_\mu \lambda_\nu (\partial^2{}_{\mu,\nu} f_i)_{x=p} + \sum_{\mu=1}^{m} \lambda'_\mu (\partial_\mu f_i)_{x=p} = 0.$$

But by definition we have $h_m(s) = s$. Thus we get $\lambda_m = h'_m(\sigma) = 1$ and $\lambda'_m = h''_m(\sigma) = 0$. Hence from (1) and (3) we obtain

$$\gamma_{i,m} = -\sum_{\mu=1}^{m-1} \{ \sum_{\nu=1}^{m-1} \lambda_\mu \lambda_\nu \partial^2{}_{\mu,\nu} \phi_i(p) + 2\lambda_\mu \partial^2{}_{\mu,m} \phi_i(p) + \lambda_\mu \gamma_{i,\mu} + \lambda'_\mu (\partial_\mu \phi_i(p) + \beta_{i,\mu} + \gamma_{i,\mu} p_\mu) \} - \partial^2{}_{m,m} \phi_i(p).$$

And by (1) and (2) we get

$$\beta_{i,m} = -\sum_{\mu=1}^{m-1} \lambda_\mu (\partial_\mu \phi_i(p) + \beta_{i,\mu} + \gamma_{i,\mu} p_\mu) - \gamma_{i,m} p_m - \partial_m \phi_i(p).$$

Thus (α, β, γ) can be represented by continuously differentiable functions ot $m + 2m^2 - 1 = q - 1$ arguments t, p_j, λ_μ, λ'_μ, $\beta_{i,\mu}$, $\gamma_{i,\mu}$ $(i, j = 1, \cdots, m,$ $\mu = 1, \cdots, m - 1)$. Therefore by Theorem 2 the set of points (α, β, γ) such that ii) holds but not iii) has measure zero in $E^q{}_{(\alpha,\beta,\gamma)}$.

In conclusion we have that, under the conditions $|\alpha_i| < \delta$, $|\beta_{i,\mu}| < \delta$, $|\gamma_{i,\mu}| < \delta$, except for a set of points (α, β, γ) of measure zero the statements ii) and iii) are satisfied by $f(x)$ defined by (1).

3. Proof of the continuity of the degree of mapping.

THEOREM 5. *Let $x' = F_t(x)$ be a regular mapping of a bounded open set G in E^m into E^m and $F_t(x)$ be a continuous function of (t, x) for $0 \leqq t \leqq 1$, $x \in \bar{G}$. If $a(t)$ $(\in E^m)$ is continuous and $a(t) \bar{\in} F_t(\bar{G} - G)$ for $0 \leqq t \leqq 1$, then $A[a(t), G, F_t]$ is constant for $0 \leqq t \leqq 1$ (except for those values of t such that $a(t)$ is on the crease of F_t).*

Proof. For the case $m=1$ the proof is not difficult, because one can approximate the mapping function including its derivative by a function which has only a finite number of critical points. Therefore we assume that the theorem holds for $m=1$, and we suppose that $m>1$.

At first let us consider the case that $F_t(x)$ does not depend on t, and $a(t)$ is continuously differentiable. It is sufficient to prove $A[a(0),G,F] = A[a(1),G,F]$. Let ϵ_0 be the distance between the curve $x=a(t)$ and $F(\bar{G}-G)$. Then by Theorem 3 and Theorem 4 there exists a regular mapping $x'=f(x)$ of G into E^m with the following properties:

i) $|f(x)-F(x)|<\epsilon_0$ for $x \,\varepsilon\, \bar{G}$.

ii) $A[a(i),G,f]=A[a(i),G,F]$ $(i=0,1)$.

iii) Rank of $(\partial_x f) \geqq m-1$ for any x such that $f(x)=a(t)$ $(0 \leqq t \leqq 1)$.

iv) For any fixed t in $0 \leqq t \leqq 1$ the set of the roots of $f(x)=a(t)$ has no accumulation point in G.

Now we take any fixed value of t in $0 \leqq t \leqq 1$ and put $a(t)=a$. Then from i) and the definition of ϵ_0 we have $a \,\bar{\varepsilon}\, f(\bar{G}-G)$. And from iv) the equation $f(x)=a$ has only a finite number of roots $c^1,\cdots,c^k$. We can find a positive number Δ such that, if U_ν is any neighborhood of c^ν with diameter smaller than Δ, the U_ν all lie in G and have no common points with one another. Since $a \,\bar{\varepsilon}\, f(\bar{G}-U_1-\cdots-U_k)$, all roots of $f(x)=a'$ lie in $U_1 \cup \cdots \cup U_k$ for any point a' with

(1) $$|a-a'|<\operatorname{dist}[a,f(\bar{G}-U_1-\cdots-U_k)]=\Delta',$$

and

(2) $$A[a',G,f]=\sum_{\nu=1}^{k} A[a',U_\nu,f],$$

if a' is not on the crease of f.

Let us consider each $A[a',U_\nu,f]$ separately. From iii) there exist $m-1$ components of f, say $f_1,\cdots,f_{l-1},f_{l+1},\cdots,f_m$, such that

$$D(f_1,\cdots,f_{l-1},f_{l+1},\cdots,f_m/x_1,\cdots,x_{m-1}) \neq 0 \text{ at } x=c^\nu.$$

If we put

(3) $$y_i=f_i(x) \text{ for } i \neq l \text{ and } y_l=x_m,$$

then $D(y/x)\neq 0$ at $x=c^\nu$. By a theorem of implicit functions we can find a positive number δ, such that the neighborhood V_ν of the point $(a_1,\cdots,a_{l-1},c_m{}^\nu,a_{l+1},\cdots,a_m)$ in $E_{(y)}{}^m$ defined by

$$V_\nu\colon \alpha_i<y_i<\beta_i \qquad (i=1,\cdots,m),$$

where $\alpha_i = a_i - \delta$, $\beta_i = a_i + \delta$ for $i \neq l$ and $\alpha_l = c_m{}^\nu - \delta$, $\beta_l = c_m{}^\nu + \delta$, corresponds homeomorphically to a neighborhood U_ν of c^ν in $E_{(x)}{}^m$ whose diameter is smaller than Δ. Let us denote its inverse mapping by $x = \phi(y)$, and put $f(\phi(y)) = g(y)$. We can suppose that the correspondence is continuous on the boundary $\bar{U}_\nu - U_\nu = \phi(\bar{V}_\nu - V_\nu)$ and that $D(\phi/y) \neq 0$ in V_ν. If we put $\operatorname{sign}\{D(\phi/y)\} = e_\nu$, which is constant in V_ν, we have by the definition of the degree of mapping

$$(4) \qquad A[a', U_\nu, f] = e_\nu A[a', V_\nu, g].$$

From $x' = f(x) = g(y)$ and (3) we have $g_i(y) = y_i$ for $i \neq l$. Therefore $D(g/y) = \partial_l g_l(y)$. Consequently we get

$$(5) \qquad A[a', V_\nu, g] = A[a'_l, (\alpha_l, \beta_l), g_{l,(a')}],$$

where the right-hand side means the degree of mapping by the one-dimensional transformation $x'_l = g_l(a'_1, \cdots, a'_{l-1}, y_l, a'_{l+1}, \cdots, a'_m) = g_{l,(a')}(y_l)$. Let a'' be any point in $|a'' - a| < \Delta'$ not on the crease of f; then the segment $a(s) = (1-s)a' + sa''$ $(0 \leqq s \leqq 1)$ lies in U_ν and $a(s) \mathrel{\bar{\varepsilon}} g(\bar{V}_\nu - V_\nu)$. Therefore by the hypothesis $A[a_l(s), (\alpha_l, \beta_l), g_{l,[s]}]$ is constant for $0 \leqq s \leqq 1$, where $g_{l,[s]}$ means the one-dimensional transformation $x'_l = g_{l(a(s))}(y_l)$. Thus by (5) we have

$$(6) \qquad A[a', V_\nu, g] = A[a'', V_\nu, g].$$

From (2), (4), and (6) we get

$$A[a', G, f] = A[a'', G, f].$$

This shows that $A[a', G, f]$ does not depend on a', when a' is sufficiently near a point of the curve $x = a(t)$ $(0 \leqq t \leqq 1)$. By using Borel's covering theorem for the curve $x = a(t)$ we obtain $A[a(0), G, f] = A[a(1), G, f]$. And by ii) we get $A[a(0), G, F] = A[a(1), G, F]$. This completes the first part of the proof.

Now we shall consider the case that $F_t(x)$ depends on t. Let τ be any value of t in $0 \leqq t \leqq 1$ and put $\operatorname{dist}[a(\tau), F_\tau(\bar{G} - G)] = \Delta$. Then there exists a $\delta > 0$, such that for $|t - \tau| < \delta$, $x \mathrel{\varepsilon} \bar{G}$, the inequalities

$$|F_t(x) - F_\tau(x)| < \Delta/2 \quad \text{and} \quad |a(t) - a(\tau)| < \Delta/2$$

hold. Let t_0, t_1 be any two values of t in $|t - \tau| < \delta$ and put

$$a[s] = (1-s)a(t_0) + sa(t_1), \qquad F_{(s)}(x) = (1-s)F_{t_0}(x) + sF_{t_1}(x).$$

Then we have $a[s] \mathrel{\bar{\varepsilon}} F_{(s)}(\bar{G} - G)$ for $0 \leqq s \leqq 1$. Let H be the open set in $E_{(s,x)}{}^{m+1}$ defined by $-\epsilon < s < 1 + \epsilon$ $(\epsilon > 0)$, $x \mathrel{\varepsilon} G$, and $(s', x') = \Phi(s, x)$

be the transformation in $E_{(s,x)}{}^{m+1}$ defined by $s' = s$, $x' = F_{(s)}(x)$. Then Φ is a regular mapping of H into $E_{(s,x)}{}^{m+1}$, and $a^*(t) = (t, a[t])$ does not touch $\Phi(\bar{H} - H)$ for $0 \leqq t \leqq 1$. Therefore $A^{m+1}[a^*(t), H, \Phi]$ is constant for $0 \leqq t \leqq 1$ by the first part of the proof. The superscript of A denotes the dimension of the space in which the mapping is considered. But one can easily prove also that

$$A^{m+1}[a^*(t), H, \Phi] = A^m[a[t], G, F_{(t)}].$$

Thus we get $A[a(t_0), G, F_{t_0}] = A[a(t_1), G, F_{t_1}]$. As t_0 and t_1 are arbitrary values in $|t - \tau| < \delta(\tau)$ and τ in $0 \leqq t \leqq 1$, then by the covering theorem of Borel we obtain the constancy of $A[a(t), G, F_t]$ for $0 \leqq t \leqq 1$. Thus the proof is complete.

Remark. Let $x' = F(x)$ be a regular mapping of a bounded open set G in E^m and a_1 and a_2 be any two points such that $|a_i - a| < \operatorname{dist}[a, F(\bar{G} - G)]$. Then $A[a_1, G, F] = A[a_2, G, F]$, if a_1 and a_2 are not on the crease of F. From this we can define $A[a, G, F]$, *even when a is on the crease of F, only provided a is not in $F(\bar{G} - G)$.* For, in any small neighborhood V of a there is a point a' not on the crease of F; and $A[a', G, F]$ is independent of a', when a' is in a sufficiently small neighborhood V of a. Therefore *Theorem 5 holds even when $a(t)$ is on the crease of F_t.*

4. The degree of mapping of continuous transformations. Now we shall extend the definition of degree of mapping to any continuous mapping of an open set in E^m. First let us consider a bounded open set and a mapping which is continuous on its closure. *Let G be a bounded open set in E^m, $x' = f(x)$ a continuous mapping of $\bar{G}$ into E^m and a a point not on $f(\bar{G} - G)$. Then $A[a, G, F]$ has the same value for any regular mapping $x' = F(x)$ of G into E^m, if F satisfies the condition*

$$(1) \qquad |F(x) - f(x)| < \operatorname{dist}[a, f(\bar{G} - G)] \qquad \text{for } x \in \bar{G}.$$

Thus we define the degree of mapping of G at a (by f) by

$$A[a, G, f] = A[a, G, F].$$

In order to prove that $A[a, G, F]$ has the same value for any regular mapping F of G under the conidtion (1), let us take two regular mappings F_0 and F_1 which satisfy the condition (1), and put $F_t(x) = (1 - t)F_0(x) + tF_1(x)$. Then $F_t(x)$ satisfies the condition (1) for $0 \leqq t \leqq 1$, and hence $a \,\bar{\varepsilon}\, F_t(\bar{G} - G)$. Thus, by Theorem 5, $A[a, G, F_t]$ is constant for $0 \leqq t \leqq 1$, and we obtain $A[a, G, F_0] = A[a, G, F_1]$.

The degree of mapping thus defined also satisfies the properties i), ii), iii), and iv) in § 1, as one can prove easily, when $f(x)$ is continuous in $\bar{G}$.

Next let us consider a more general case: Let G be an open set in E^m and $x' = f(x)$ be a continuous mapping of G into E^m. *Let $(f; G)$ be the set of all accumulation points of $\{f(p_n)\}$ where $\{p_n\}$ is any sequence of points of G having no accumulation point in G.* If G is bounded and f is continuous on $\bar{G}$ (the closure of G), then $(f; G) = f(\bar{G} - G)$. If $a \bar{\varepsilon} (f; G)$, then the set R of all roots of the equation $f(x) = a$ in G is a bounded closed set in E^m. Let G_i $(i = 1, 2)$ be any two bounded open sets with the properties $R \subset G_i$ and $\bar{G}_i \subset G$. Then f is continuous on $\bar{G}_i$ and it is easy to see that

$$A[a, G_1, f] = A[a, G_2, f].$$

Thus *we define $A[a, G, f]$ by $A[a, G, f] = A[a, G_1, f]$ where G_1 is any bounded open set with the properties $R \subset G_1$ and $\bar{G}_1 \subset G$, where R is the set of all roots of $f(x) = a$ in G.*

The properties i), ii), and iii) in § 1 hold also for the degree of mapping of this kind, if one replaces $f(\bar{G} - G)$ by $(f; G)$. To extend the statement iv) to this case, we define $(f_t; G, \tau)$ as follows:

Let f_t be a continuous mapping of G into E^m depending on the parameter t, $0 \leqq t \leqq 1$. Then $(f_t; G, \tau)$ means the set of all accumulation points of $\{f_{t_n}(p_n)\}$ where $\lim t_n = \tau$ and $\{p_n\}$ is any sequence of points of G having no accumulation point in G. If G is bounded and f_t is a continuous function of (t, x) for $0 \leqq t \leqq 1$, $x \varepsilon \bar{G}$, then we have $(f_t; G, \tau) = f_\tau(\bar{G} - G)$ for any $0 \leqq \tau \leqq 1$.

To prove the statement iv) where $f_t(\bar{G} - G)$ is replaced by $(f_t; G, \tau)$, we use the following lemma which one can prove easily:

If $a \bar{\varepsilon} (f_t; G, \tau)$, then there exists a neighborhood V of a, a positive number δ and a bounded open set G_1, such that $\bar{G}_1 \subset G$ and $V \cap f_t(G - G_1)$ is empty for $| t - \tau | < \delta$.

We can take δ so small that $a(t) \varepsilon V$ also holds for $| t - \tau | < \delta$. Thus we have for $| t - \tau | < \delta$,

$$A[a(t), G, f_t] = A[a(t), G_1, f_t].$$

As $\bar{G}_1 - G_1 \subset G - G_1$, we have also $a(t) \bar{\varepsilon} f_t(\bar{G}_1 - G_1)$ for $| t - \tau | < \delta$. Then $A[a(t), G_1, f_t]$ is constant for $| t - \tau | < \delta$. Consequently $A[a(t), G, f_t]$ is constant for $| t - \tau | < \delta$. By Borel's covering theorem applied to the closed interval $0 \leqq t \leqq 1$ *we obtain statement* iv), *replacing $a(t) \bar{\varepsilon} f_t(\bar{G} - G)$ by $a(\tau) \bar{\varepsilon} (f_t; G, \tau)$.*

5. The product theorem.

THEOREM 6. *Let G be an open set in E^m and $x' = f(x)$ a continuous mapping of G into E^m. Let H be an open set containing $f(G)$ and H_i $(i = 1, 2, \cdots)$ the components of the open set $H - (f; G)$. Let $x' = \phi(x)$ be a continuous mapping of H into E^m and a a point of E^m outside of $(\phi f; G) \cup (\phi; H)$. Then we have*

$$A[a, G, \phi f] = \sum_i A[a, H_i, \phi] A[b_i, G, f], \tag{1}$$

where b_i is an arbitrary point of H_i.

Proof. $A[p, G, f]$ is constant when p varies on a domain H_i. Let D_k be the sum of those domains H_i such that $A[p, G, f] = k$ for $p \varepsilon H_i$. Then D_k is the set of all points $p \varepsilon H$ such that $A[p, G, f] = k$.

Now let R be the set of all roots of $\phi(x) = a$. Then since R is compact and $R \cap (f; G)$ is empty (because $a \bar{\varepsilon} \phi(H \cap (f; G)) \subset (\phi f; G)$), R can be covered by a finite number of H_i. Thus we get by property iii),

$$\sum_i A[a, H_i, \phi] A[b_i, G, f] = \sum_k A[a, D_k, \phi] k. \tag{2}$$

Therefore it suffices to prove the relation

$$A[a, G, \phi f] = \sum_k A[a, D_k, \phi] k. \tag{3}$$

The relation (3) is true, if G and H are bounded, f and ϕ are regular mappings, and a is not on the crease of ϕf, as one can prove easily from the definition of the degree of mapping.

In order to discuss the general case we first assume that G and H are bounded, that f and ϕ are continuous on $\bar{G}$ and $\bar{H}$ respectively and that $f(\bar{G}) \subset H$. Let us take regular transformations f^* and ϕ^* sufficiently near to f and ϕ respectively such that a is not on the crease of $\phi^* f^*$:

$$| f^*(x) - f(x) | < \operatorname{dist}[R, f(\bar{G} - G)] \text{ for } x \varepsilon \bar{G},$$

$$| \phi^*(x) - \phi(x) | < \operatorname{dist}[a, \phi f^*(\bar{G} - G) \cup \phi(\bar{H} - H)] \text{ for } x \varepsilon \bar{H},$$

and

$$| \phi^* f^*(x) - \phi f(x) | < \operatorname{dist}[a, \phi f(\bar{G} - G)] \text{ for } x \varepsilon \bar{G}.$$

Thus, if we denote by D^*_k the set of all points $p \varepsilon H$ such that $A[p, G, f^*] = k$, we have

$$\operatorname{dist}[a, \phi f^*(\bar{G} - G) \cup \phi(\bar{H} - H)] \leqq \operatorname{dist}[a, \phi(\bar{D}^*_k - D^*_k)];$$

$$A[a, D^*_k, \phi^*] = A[a, D^*_k, \phi] = A[a, D_k, \phi], \tag{4}$$

since $D^*_k \cap R = D_k \cap R$ (because $A[p, G, f^*] = A[p, G, f]$ for $p \varepsilon R$); and

$$(5) \qquad A[a, G, \phi^* f^*] = A[a, G, \phi f].$$

But, as we have already said, we have

$$(6) \qquad A[a, G, \phi^* f^*] = \sum_k [a, D^*_k, \phi^*]k.$$

Thus from (4), (5), and (6) we get

$$(7) \qquad A[a, G, \phi f] = \sum_k A[a, D_k, \phi]k.$$

Next we discuss the general case, when G and H are not necessarily bounded and f and ϕ are continuous respectively on G and H. Let $\{G_n\}$ and $\{H_n\}$ be sequences of bounded open sets such that $\bar{G}_n \subset G_{n+1}$, $\lim G_n = G$, $f(\bar{G}_n) \subset H_n$, $\bar{H}_n \subset H_{n+1}$, $\lim H_n = H$, and $D_k{}^n$ the set of all points $p \varepsilon H_n$ such that $A[p, G_n, f] = k$. Then $f(\bar{G}_n - G_n)$ approaches $(f; G)$ as $n \to \infty$, and $\underline{\lim}\, D_k{}^n \supseteq D_k$ (as point sets). Since R can be covered by a finite number of D_k, for example by $D_{-l}, \cdots, D_l$, and each $D_k \cap R$ is compact, it follows that R can be covered by $D_{-l}{}^n, \cdots, D_l{}^n$ for a sufficiently large n. Thus we have

$$A[a, G_n, \phi f] = A[a, G, \phi f] \text{ and } A[a, D_k{}^n, \phi] = A[a, D_k, \phi]$$

for a certain fixed n and all k. Therefore (7) holds for the general case. Thus from (2) and (7) we obtain (1).

OSAKA UNIVERSITY,
OSAKA, JAPAN.

DEGREE OF MAPPING IN CONVEX LINEAR TOPOLOGICAL SPACES.*

By Mitio Nagumo.

J. Leray and J. Schauder have developed a theory of the degree of mapping for a completely continuous movement [1] in a Banach space (1934) [1]. Before this, Schauder gave a theorem on the invariance of domain under a one-one completely continuous movement in a weakly compact Banach space (1929) [6], (1932) [7]. These works contain important applications to the theory of partial differential equations of elliptic type. Leray also gave a product theorem concerning the degree of mapping for the composition of completely continuous movements in general Banach spaces and extended the theorem of Schauder on the invariance of domain to the case of general Banach spaces (1935) [2]. However, the treatments [1] and [2] seem not to be complete in detail. In this paper, following Leray and Schauder, I wish to establish a theory of the degree of mapping in the case of convex linear topological spaces and prove a theorem on the invariance of domain in the case of convex linear metric complete spaces.[2]

I am much obliged to the referee of this Journal for several corrections of my manuscript and for references to the literature.

1. Preliminary notions.

1.1. First we shall explain the notion of the degree of mapping in a finite dimensional Euclidean space. Let G be an open set in an m-dimensional Euclidean space E^m and $x' = f(x) = \{f_i(x) \mid i = 1, \cdots, m\}$ $(x = (x_1, \cdots, x_m))$ be a continuous mapping of $\bar{G}$ (the closure of G) into E^m such that $f(x) - x$ is bounded on $\bar{G}$. Let a be a point of E^m *not in* $f(\bar{G} - G)$ (the image of $\bar{G} - G$ by f). Then there will be determined an integer $A[a, G, f]$, called the *degree of mapping* of G at a by f, with the following properties:

i) *If f is the identical mapping $f(x) = x$, then*

$A[a, G, f] = 1$ *when* $a \varepsilon G$, $A[a, G, f] = 0$ *when* $a \bar{\varepsilon} \bar{G}$.[3]

---

* Received March 6, 1950; revised September 4, 1950.

[1] I. e., a transformation of the form $x' = x + f(x)$, where f is completely continuous.

[2] I am also informed about the work of E. Rothe on the same subject [5], but I must sorrily confess that I have had no chance to see it.

[3] $\bar{\varepsilon}$ means the negation of ε.

ii) *If* $A[a, G, f] \neq 0$, *then there exists a point* $x \varepsilon G$ *such that* $f(x) = a$.

iii) *If* G *is divided into open sets* $G_1, \cdots, G_k$, *i. e.* $G \supset \bigcup_{i=1}^{k} G_i$, $\bar{G} = \bigcup_{i=1}^{k} \bar{G}_i$, *and* $G_i \cap G_j = 0$ *(empty set)* $(i \neq j)$, *and if* $a \bar{\varepsilon} f(\bar{G}_i - G_i)$ *for any* i, *then*

$$A[a, G, f] = \sum_{i=1}^{k} A[a, G_i, f].$$

iv) *If* $f_t(x) - x$ *is a bounded continuous function of* (t, x) *for* $0 \leqq t \leqq 1$, $x \varepsilon \bar{G}$, *if* $a(t) (\varepsilon E^m)$ *is continuous, and if* $a(t) \bar{\varepsilon} f_t(\bar{G} - G)$ *for all* t *in* $0 \leqq t \leqq 1$, *then* $A[a(t), G, f_t]$ *is constant for* $0 \leqq t \leqq 1$.

v) *Let* a *be a point not on* $f(\bar{G} - G)$, X *the set of all roots of the equation* $f(x) = a$ *in* G, *and* G_0 *any open set such that* $X \subset G_0 \subset G$, *then*

$$A[a, G_0, f] = A[a, G, f].$$

Remark. v) follows from ii) and iii).

The existence and uniqueness of $A[a, G, f]$ satisfying the above conditions can be verified, if we use simplicial mappings for approximations of f (at first for bounded G and then for general G). But I may refer to [3] in which the existence of $A[a, G, f]$ is given, based on infinitesimal analysis but free from the notion of simplicial mapping. (It will be easily seen that $f(\bar{G} - G) = (f; G)$ and $f_\tau(\bar{G} - G) = (f_t; G, \tau)$ in the notation of [3].)

From iv) we have

COROLLARY. *If* f_0 *and* f_1 *are continuous transformations of* $\bar{G}$ *into* E^m *such that* $|f_1(x) - f_0(x)| < \operatorname{dist}(a, f_0(\bar{G} - G))$,[4] *then*

[4] $|x| = (\sum_{i=1}^{m} x_i^2)^{1/2}$ for $x \varepsilon E^m$.

$$A[a, G, f_1] = A[a, G, f_0].$$

Proof. Put $f_t(x) = (1 - t) f_0(x) + t f_1(x)$ and apply iv).

1.2. A linear topological space is a linear set on which is imposed a topology in such a fashion that the postulated operations of addition of elements and multiplication of elements by real numbers are continuous in the topology. Cf. [4] and [9]. It suffices to give the system $\mathfrak{U}$ of neighborhoods of the origin. The system of neighborhoods of an arbitrary point a consists of the neighborhoods of the form $U(a) = a + U (U \varepsilon \mathfrak{U})$. A linear topological space E is called *convex*, if the topology of E is defined by means

of a system of neighborhoods which are convex open sets, i. e., if $U \varepsilon \mathfrak{U}$, $\alpha, \beta > 0$, $\alpha + \beta = 1$, then $\alpha U + \beta U = U$.[5] We assume also, without loss of generality, the symmetry of U, $-U = U$, for $U \varepsilon \mathfrak{U}$. Hence $b \varepsilon U(a)$ is equivalent with $a \varepsilon U(b)$.

In a convex linear topological space E the *pseudonorm* $\| x \|_U (U \varepsilon \mathfrak{U})$ is defined for every $x \varepsilon E$ as follows (cf. [4]):

$$\| x \|_U = \text{g. l. b. } \textit{of } \alpha > 0 \textit{ such that } x \varepsilon \alpha U.$$

Then $\| x \|_U$ has the following properties:

(a) $\| x \|_U \geqq 0$,

(b) $\| x + y \|_U \leqq \| x \|_U + \| y \|_U$,

(c) $\| \alpha x \|_U = | \alpha | \, \| x \|_U$ *for any real* α,

(d) $\| x \|_U$ *is a continuous function of* x,

(e) $\| x \|_U < 1$ *is equivalent with* $x \varepsilon U$.

In this paper *we denote by E always a convex linear topological space*, and assume $U(a)$, $\mathfrak{U}$ have the meanings mentioned above.

2. Auxiliary theorems. Before we proceed to the definition of the degree of mapping in E we give some auxiliary theorems. A transformation f of M into E is called *completely continuous* on M if f is continuous on M and the image $f(M)$ of M is a subset of a compact set in E.[6] In this paper we denote by $\mathsf{T}f$ the transformation

$$\mathsf{T}f(x) = x + f(x)$$

for any transformation f in E.

THEOREM 1. *Let M be a closed set in E, and f be completely continuous on M. Then* $\mathsf{T}f(M)$ *is also closed in E.*

Proof. We have to prove that for any point $a \bar{\varepsilon} \mathsf{T}f(M)$ there exists a $U \varepsilon \mathfrak{U}$ such that $U(a) \cap \mathsf{T}f(M) = 0$.

For any $p \varepsilon E$ there exists a $U_p \varepsilon \mathfrak{U}$ such that:

$$U_p(p) \cap M = 0 \text{ if } p \bar{\varepsilon} M, \tag{1}$$

$$U_p(p) \cap (a - f(U_p(p))) = 0 \text{ if } p \varepsilon M, \tag{2}$$

---

[5] By αS (α a real number), we denote the set of all points αx, where $x \varepsilon S$, and by $S_1 + S_2$ the set of all points $x_1 + x_2$, where $x_i \varepsilon S_i$.

[6] This definition is more restricted than the ordinary one, when M is not bounded. We use the term "compact" in the sense of "bicompact."

since $p \neq a - f(p)$ and f is continuous on M. The set $a - f(M)$ is contained in a compact set K, and there exist a finite number of points $p_i \varepsilon K$ $(i = 1, \cdots, k)$ such that

$$(3) \qquad \bigcup_{i=1}^{k} 2^{-1} U_{p_i}(p_i) \supset K.$$

Then we shall have $U_0(a) \cap \mathsf{T}f(M) = 0$, if we put

$$(4) \qquad U_0 = \bigcap_{i=1}^{k} 2^{-1} U_{p_i}.$$

For if $U_0(a) \cap \mathsf{T}f(M) \neq 0$, then there should exist a $p \varepsilon M$ such that $p + f(p) \varepsilon U_0(a)$, hence $p \varepsilon U_0(a - f(p)) \subset U_0(K)$. Then by (3) and (4) there exists an i such that $a - f(p) \varepsilon U_{p_i}(p_i)$ and $p \varepsilon U_{p_i}(p_i)$. If $p_i \bar{\varepsilon} M$, then by (1) $U_{p_i}(p_i) \cap M = 0$, which contradicts $p \varepsilon M$ and $p \varepsilon U_{p_i}(p_i)$. If $p_i \varepsilon M$, then by (2) $U_{p_i}(p_i) \cap (a - f(U_{p_i}(p_i))) = 0$, which contradicts $a - f(p) \varepsilon U_{p_i}(p_i)$ and $a - f(p) \varepsilon (a - f(U_{p_i}(p_i)))$. Thus the proof is complete.

THEOREM 2. *Let K be a compact set in E. Then for any $U \varepsilon \mathfrak{U}$ there exists a finite-dimensional linear manifold E^m in E and a continuous transformation S of K into E^m such that $S(x) - x \varepsilon U$ for $x \varepsilon K$.*

Proof. By the hypothesis there exist a finite number of points $p_i \varepsilon K$ $(i = 1, \cdots, k)$ such that $K \subset \bigcup_{i=1}^{k} U(p_i)$. Let E^m be the linear manifold spanned by $p_1, \cdots, p_k$. We put $\rho_i(x) = \mathrm{Max}\{(1 - \| x - p_i \|_U), 0\}$ and define S by

$$S(x) = (\sum_{i=1}^{k} \rho_i(x))^{-1} \sum_{i=1}^{k} \rho_i(x) p_i.$$

That S satisfies the above statement follows from the convexity of U, the continuity of $\rho_i(x)$, and the equivalence of $\rho_i(x) > 0$ with $x \varepsilon U(p_i)$.

THEOREM 3. *Let G be an open set in an l-dimensional Euclidean space E^l (l finite); f a continuous transformation of $\bar{G}$ into E^l such that $f(x) - x = \phi(x)$ is bounded on G and $\phi(\bar{G}) \subset E^m$, where E^m is an m-dimensional linear manifold in $E^l (m < l)$; and $a \bar{\varepsilon} f(\bar{G} - G)$; then*

$$A^m[a, G^m, f] = A^l[a, G, f],$$

where $G^m = G \cap E^m$ and the superscript of A means that the degree of mapping is considered in the space of the assigned dimension.

Proof. Because the degree of mapping is invariant under a linear

transformation of the coordinates, let us take the x_1-axis, $\cdots$, x_m-axis in E^m so that the transformation $x' = f(x)$ shall be of the form

$$x'_i = f_i(x_1, \cdots, x_m, \cdots, x_l) \text{ for } 1 \leqq i \leqq m,$$
$$x'_j = x_j = f_j(x) \text{ for } m+1 \leqq j \leqq l.$$

Let R be an open sphere with center at the origin and a sufficiently large radius; then, putting $G_0{}^i = R \cap G^i$, we have by v) in § 1,

$$A^i[a, G^i, f] = A^i[a, G_0{}^i, f] \qquad (i = l, m). \tag{1}$$

Let the distance between a and $f(\bar{G}_0 - G_0)$ be Δ, so that $\Delta > 0$; and let $f_i^*(x)$ be functions with continuous partial derivatives on $\bar{G}_0$ such that $\sum_{i=1}^{m} \{f_i^*(x) - f_i(x)\}^2 < \Delta^2$ and $f_j^*(x) = f_j(x) = x_j$ for $m+1 \leqq j \leqq l$. Then by the Corollary in § 1,

$$A^i[a, G_0{}^i, f] = A^i[a, G_0{}^i, f^*] \qquad (i = l, m). \tag{2}$$

Using the notation of [3] we have

$$D(f_1^*, \cdots, f_l^*/x_1, \cdots, x_l) = D(f_1^*, \cdots, f_m^*/x_1, \cdots, x_m).^7$$

Thus by the definition of $A[a, G, f^*]$ in [3] we obtain

$$A^l[a, G_0{}^l, f^*] = A^m[a, G_0{}^m, f^*],$$

since all roots of $f^*(x) = a$ are in $E^m (a_j = 0$ for $j > m)$. From this with (1) and (2) the theorem follows.

3. Definition of the degree of mapping in E. Let G be an open set in E and f a completely continuous transformation of G into E. Let a be a point not on $\mathsf{T}f(\bar{G} - G)$; then by Theorem 1 there exists a $U \varepsilon \mathfrak{U}$ such that $U(a) \cap \mathsf{T}f(\bar{G} - G) = 0$. By hypothesis there exists a compact set K such that $f(\bar{G}) \subset K$ and by Theorem 2 there exist an m-dimensional linear manifold E^m (m finite) and a continuous transformation S of K into E^m such that $a \varepsilon E^m$, $S(x) - x \varepsilon U$ for $x \varepsilon K$. Then $\mathsf{T}Sf(x) - \mathsf{T}f(x) \varepsilon U$ for $x \varepsilon \bar{G}$; hence $a \bar{\varepsilon} \mathsf{T}Sf(\bar{G} - G)$. Since a linear topological space of finite dimension is linearly homeomorphic to the Euclidean space of the same dimension (cf. [8]), $\mathsf{T}Sf$ transforms $\bar{G}^m = \bar{G} \cap E^m$ into E^m, $\mathsf{T}Sf(x) - x = Sf(x)$ is bounded ($S(K)$ is compact) on $\bar{G}^m$, and $a \bar{\varepsilon} \mathsf{T}Sf(\bar{G}^m - G^m)$ (for $\bar{G}^m - G^m \subset \bar{G} - G$), it follows that $A^m[a, G^m, \mathsf{T}Sf]$, as the degree of mapping

[7] $D(f_1, \cdots, f_k/x_1, \cdots, x_k) = \det(\partial f_i/\partial x_j)_{i,j=1,\ldots,k}$.

in E^m, has a definite meaning. Then *we define the degree of mapping* $A[a, G, \mathsf{T}f]$ *in* E *by*

$$(0) \qquad A[a, G, \mathsf{T}f] = A^m[a, G^m, \mathsf{T}Sf].$$

This definition will be legitimized by the following:

THEOREM 4. $A[a, G, \mathsf{T}f]$ *defined by* (0) *is independent of the choice of* E^m *and* S.

Proof. Let $U_i \,\varepsilon\, \mathfrak{U}$ $(i = 1, 2)$ be such that

$$(1) \qquad U_i(a) \cap \mathsf{T}f(\bar{G} - G) = 0.$$

Let E^m and E^n be linear manifolds of m respectively n dimensions in E, and S_i $(i = 1, 2)$ be continuous transformations of K into E such that

$$(2) \qquad S_i(x) - x \,\varepsilon\, U_i \text{ for } x \,\varepsilon\, K,\ a \,\varepsilon\, E^m \cap E^n,\ S_1(K) \subset E^m,\ S_2(K) \subset E^n.$$

By Theorems 1 and 2 we can select a linear manifold E^l of finite dimension l in E and a continuous transformation S_3 of K into E^l such that

$$(3) \qquad S_3(x) - x \,\varepsilon\, U_1 \cap U_2 \text{ for } x \,\varepsilon\, K \text{ and } E^m \cup E^n \subset E^l.$$

Then, putting $G^i = G \cap E^i$ $(i = l, m, n)$, we get by Theorem 3

$$(4) \qquad A^m[a, G^m, \mathsf{T}S_1f] = A^l[a, G^l, \mathsf{T}S_1f].$$

If we put $(1 - t)\mathsf{T}S_1f + t\mathsf{T}S_3f = F_t$, then by (2) and (3)

$$F_t(x) - \mathsf{T}f(x) = (1 - t)\{S_1f(x) - f(x)\} + t\{S_3f(x) - f(x)\}.$$

$$\varepsilon\, U_1 \text{ for } x \,\varepsilon\, \bar{G}, \quad 0 \leqq t \leqq 1.$$

Thus from (1) and $\bar{G}^l - G^l \subset \bar{G} - G$ it follows that

$$(5) \qquad a \,\bar{\varepsilon}\, F_t(\bar{G}^l - G^l) \text{ for } 0 \leqq t \leqq 1.$$

Since $F_t(x) - x = (1 - t)S_1f(x) + S_3f(x)$ is bounded ($S_i(K)$ are compact) for $x \,\varepsilon\, \bar{G}^l$, then by (5) and iv) in § 1, $A^l[a, G^l, F_t]$ is constant for $0 \leqq t \leqq 1$. Thus $A^l[a, G^l, \mathsf{T}S_1f] = A^l[a, G^l, \mathsf{T}S_3f]$. Hence by (4)

$$A^m[a, G^m, \mathsf{T}S_1f] = A^l[a, G^l, \mathsf{T}S_3f].$$

Similarly we get

$$A^n[a, G^n, \mathsf{T}S_2f] = A^l[a, G^l, \mathsf{T}S_3f].$$

Consequently

$$A^m[a, G^m, \mathsf{T}S_1f] = A^n[a, G^n, \mathsf{T}S_2f].$$

This completes the proof.

4. Fundamental properties of the degree of mapping in E. The fundamental properties of the degree of mapping stated in § 1 remain valid also for transformations $\mathsf{T}f$ in E. The symbols f, T, K, etc. have the same meaning as in § 3. The validity of i) is evident.

THEOREM 5. *Property* ii) *in* § 1 *is valid for the mapping* $\mathsf{T}f$ *(instead of f) in E, if* $a \bar{\varepsilon} \mathsf{T}f(\bar{G} - G)$.

Proof. If $a \bar{\varepsilon} \mathsf{T}f(G)$, then $a \bar{\varepsilon} \mathsf{T}f(\bar{G})$. Thus by Theorem 1 there exists a $U \varepsilon \mathfrak{U}$ such that

$$(1) \qquad U(a) \cap \mathsf{T}f(\bar{G}) = 0.$$

Then by Theorem 2 there exists a linear manifold of finite dimension E^m and a continuous transformation S of K into E^m such that

$$(2) \qquad S(x) - x \,\varepsilon\, U \text{ for } x \,\varepsilon\, K.$$

By the hypothesis and the definition of the degree of mapping in E,

$$0 \neq A[a, G, \mathsf{T}f] = A^m[a, G^m, \mathsf{T}Sf] \quad (G^m = G \cap E^m).$$

Thus by ii) in § 1, for the mapping $\mathsf{T}Sf$ in E^m, there exists a point $x_0 \,\varepsilon\, G^m$ such that $a = \mathsf{T}Sf(x_0)$. But by (2), since $f(x_0) \,\varepsilon\, K$, $\mathsf{T}Sf(x_0) - \mathsf{T}f(x_0) = Sf(x_0) - f(x_0) \,\varepsilon\, U$. Then $\mathsf{T}f(x_0) \,\varepsilon\, U(a)$ which contradicts (1).

THEOREM 6. *Property* iii) *in* § 1 *is valid for the mapping* $\mathsf{T}f$ *(instead of f) in E.*

Proof. Since $a \,\varepsilon\, \mathsf{T}f(\bar{G}_i - G_i)$ $(i = 1, \cdots, k)$, by Theorem 1 there exists a $V \varepsilon \mathfrak{U}$ such that

$$(1) \qquad V(a) \cap \mathsf{T}f(\bar{G}_i - G_i) = 0, \qquad V(a) \cap \mathsf{T}f(\bar{G} - G) = 0,$$

since $\bar{G} - G \subset \bigcup_{i=1}^{k} (\bar{G}_i - G_i)$. By Theorem 2 there exists a finite dimensional linear manifold E^m and a continuous transformation S of K into E^m such that $a \,\varepsilon\, E^m$ and

$$(2) \qquad S(x) - x \,\varepsilon\, V \text{ for } x \,\varepsilon\, K.$$

Then by the definition of the degree of mapping for $\mathsf{T}f$ in E,

$$(3) \qquad \begin{cases} A[a, G_i, \mathsf{T}f] = A^m[a, G_i^m, \mathsf{T}Sf] \quad (i = 1, \cdots, m), \\ A[a, G, \mathsf{T}f] = A^m[a, G^m, \mathsf{T}Sf]. \end{cases}$$

From (1) and (2) also follows

$$(4) \qquad a \,\bar{\varepsilon}\, \mathsf{T}Sf(\bar{G}_i - G_i) \cup \mathsf{T}Sf(\bar{G} - G) \qquad (i = 1, \cdots, m).$$

Let X be the set of all roots of $\mathsf{T}Sf(x) = a$ in $\bar{G}$. Then $X \subset \bar{G}^m$, because from $\mathsf{T}Sf(x) = a$ follows $x = a - Sf(x) \,\varepsilon\, E^m$. From (4) and $\bar{G} = \bigcup_{i=1}^{k} \bar{G}_i$ we obtain $X \subset \bigcup_{i=1}^{k} G_i^m \subset G^m$. Thus by v) for $\mathsf{T}Sf$ (instead of f) in E^m,

$$(5) \qquad A^m[a, G^m, \mathsf{T}Sf] = A^m[a, \bigcup_{i=1}^{k} G_i^m, \mathsf{T}Sf].$$

Since $G_i^m \cap G_j^m = 0$ $(i \neq j)$, we get by ii) in § 1 for $\mathsf{T}Sf$ (instead of f) in E^m

$$(6) \qquad A^m[a, \bigcup_{i=1}^{k} G_i^m, \mathsf{T}Sf] = \sum_{i=1}^{k} A^m[a, G_i^m, \mathsf{T}Sf].$$

From (3), (5) and (6) we obtain

$$A[a, G, \mathsf{T}f] = \sum_{i=1}^{k} A[a, G_i, \mathsf{T}f].$$

COROLLARY 1. *Property* v) *in* § 1 *is valid for* $\mathsf{T}f$ (*instead of* f) *in* E.

Proof. Put $G - \bar{G}_0 = G$. Then $G \supset G_0 \cup G_1$, $\bar{G} = \bar{G}_0 \cup \bar{G}_1$, $G_0 \cap G_1 = 0$. Apply Theorem 6 and the relation $A[a, G_1, \mathsf{T}f] = 0$ (by Thoerm 5).

Property iv) in § 1 becomes as follows:

THEOREM 7. *Let f_t be a transformation of $\bar{G}$ into E such that $f_t(x)$ is a continuous function of (t, x) for $0 \leqq t \leqq 1$, $x \,\varepsilon\, \bar{G}$ and is always contained in a compact set K.*[8] *If $a(t)\,(\varepsilon\, E)$ is continuous for $0 \leqq t \leqq 1$ and $a(t) \,\bar{\varepsilon}\, \mathsf{T}f_t(\bar{G} - G)$ for $0 \leqq t \leqq 1$, then $A[a(t), G, \mathsf{T}f_t]$ is constant for $0 \leqq t \leqq 1$.*

To prove this we use the following

LEMMA 1. *Let K_i $(i = 1, 2)$ be compact sets in E; then $K_1 + K_2$*[5] *is also compact. Let K be a compact set in E; then the set $\hat{K}$ of all points tx such that $0 \leqq t \leqq 1$, $x \,\varepsilon\, K$, is also compact.*

Proof. $K_1 \times K_2$ is a compact set in $E \times E$; then by the continuous mapping $x' = x_1 + x_2 (x_i \,\varepsilon\, K_i)$, $K_1 \times K_2$ goes onto the compact set $K_1 + K_2$

---

[8] This condition is weaker than that $f_t(x)$ is completely continuous for $x \,\varepsilon\, G$ and uniformly continuous in t for $x \,\varepsilon\, \bar{G}$, $0 \leqq t \leqq 1$.

in E. Similarly $K \times \langle 0, 1\rangle$ is a compact set in $E \times E^1$, which is mapped by $x' = tx$ onto the compact set $\hat{K}$ in E.

Proof of Theorem 7. First we assume that $a(t)$ is constant and put $a(t) = a$. Let τ be any fixed value of t from $0 \leqq t \leqq 1$. Consider the product space $E^* = E \times E^1$ (E^1 is the 1-dimensional space $-\infty < t < +\infty$), which is also a convex linear topological space with a system of neighborhoods of the origin, $\mathfrak{U}^* = \{U \times (-\delta, \delta) \mid U \varepsilon \mathfrak{U}, \delta > 0\}$. Let G^* be the open set in E^* defined by $G^* = G \times (-\infty, +\infty)$. Let us define the transformation f^* in E^* by $f^*(x, t) = (f_{\langle t\rangle}(x), 0)$, where $\langle t\rangle = 0$ for $t < 0$, $\langle t\rangle = t$ for $0 \leqq t \leqq 1$, $\langle t\rangle = 1$ for $t > 1$. Then $f^*(\bar{G}^*) \subset (K, 0) = K^*$ (a compact set in E^*). Since $\bar{G}^* - G^* = (\bar{G} - G) \times (-\infty, +\infty)$, then $(a, \tau) \bar{\varepsilon} \mathsf{T}f^*(\bar{G}^* - G^*)$. Thus by Theorem 1 there exists a $U^* \subset \mathfrak{U}^*$ such that

$$U^*(a, \tau) \cap \mathsf{T}f^*(\bar{G}^* - G^*) = 0.$$

This means that there exists a $\delta > 0$ and a $U \varepsilon \mathfrak{U}$ such that

(1) $$U(a) \cap \mathsf{T}f_t(\bar{G} - G) = 0 \text{ for } |t - \tau| < \delta.$$

By Theorem 2 there exists a finite-dimensional linear manifold E^m and a continuous transformation S of K into E^m such that $a \varepsilon E^m$ and $S(x) - x \varepsilon U$ for $x \varepsilon K$. Then by (1) and the definition of the degree of mapping for $\mathsf{T}f_t$ in E,

(2) $$A[a, G, \mathsf{T}f_t] = A^m[a, G^m, \mathsf{T}Sf_t] \text{ for } |t - \tau| < \delta.$$

From (1) and $\mathsf{T}f_t(x) - \mathsf{T}Sf_t(x) \varepsilon U$ follows $a \bar{\varepsilon} \mathsf{T}Sf_t(\bar{G}^m - G^m)$ for $|t - \tau| < \delta$. Thus by iv) in § 1, $A^m[a, G^m, \mathsf{T}Sf_t]$, and hence by (2) $A[a, G, \mathsf{T}f_t]$, is constant for $|t - \tau| < \delta$. Then by using the covering theorem for the closed interval $0 \leqq t \leqq 1$ we obtain the theorem for constant $a(t)$.

The constancy of $a(t)$ can be dropped, if we consider $f_t(x) - a(t)$ instead of $f_t(x)$, since the set $K + \{-a(t) \mid 0 \leqq t \leqq 1\}$ is also compact by Lemma 1. For, $A[a, G, \mathsf{T}f] = A[0, G, \mathsf{T}f - a]$ if $a \bar{\varepsilon} \mathsf{T}f(\bar{G} - G)$.[9]

COROLLARY 2. *Let f be a completely continuous transformation of $\bar{G}$ into E, where G is an open set in E. If a is a point of E and there exists*

[9] This relation is valid if E is finite-dimensional (put $F_t = Tf - ta$, $a(t) = (1 - t)a$ and apply iv) in §1). For, the general case we have by definition $A[a, G, Tf] = A^m[a, G^m, TSf]$, $A[0, G, Tf - a] = A^m[0, G^m, TSf - a]$, where A^m, G^m and S have the same meaning as in §2. Hence the relation holds for general E.

a $U \varepsilon \mathfrak{U}$ such that $U(a) \cap \mathsf{T}f(\bar{G}-G)=0$, then for every point $a' \varepsilon U(a)$, $A[a', G, \mathsf{T}f]=A[a, G, \mathsf{T}f]$.

Proof. Put $a(t)=(1-t)a+ta'$ and apply Theorem 7.

COROLLARY 3. *Let f, G, a and U satisfy the same condition as in Corolary 2. Then for every completely continuous transformation f' of $\bar{G}$ into E such that*

$$f'(x)-f(x) \varepsilon U \text{ for } x \varepsilon \bar{G},$$

$$A[a, G, \mathsf{T}f']=A[a, G, \mathsf{T}f].$$

Proof. Put $f_t=(1-t)f+tf'$ and apply Theorem 7. $f_t(\bar{G})$ is contained in a definite compact set for $0 \leqq t \leqq 1$ by Lemma 1.

5. Product theorem. For the composition of two transformations in a finite dimensional E^m the following holds:

THEOREM 8. *Let G be an open set in E^m and f a continuous mapping of $\bar{G}$ into E^m such that $f(x)-x$ is bounded on G. Let H be an open set containing $f(\bar{G})$ and H_i $(i=1, 2, \cdots)$ the components of the open set $H-f(\bar{G}-G)$. Let ϕ be a continuous mapping of $\bar{H}$ into E^m such that $\phi(x)-x$ is bounded, and a a point of E^m such that $a \bar{\varepsilon} \phi f(\bar{G}-G) \cup \phi(\bar{H}-H)$. Then we have*

$$A[a, G, \phi f]=\sum_i A[a, H_i, \phi] \cdot A[b_i, G, f],$$

where b_i is an arbitrary point of H_i.

Proof. Cf. [3].

This theorem can also be extended to the general case of E as follows:

THEOREM 9. *Let G be an open set in E and f a completely continuous mapping of $\bar{G}$ into E. Let H be an open set containing $\mathsf{T}f(\bar{G})$ and H_i the components of the open set $H-\mathsf{T}f(\bar{G}-G)$. Let ϕ be completely continuous mapping of $\bar{H}$ into E and a a point of E such that*

$$a \bar{\varepsilon} \mathsf{T}\phi * f(\bar{G}-G) \cup \mathsf{T}\phi(\bar{H}-H),$$

where

$$\mathsf{T}\phi * f=\mathsf{T}\phi \mathsf{T}f \cdot (\phi * f=f+\phi+\phi \mathsf{T}f).$$

Then we have

$$A[a, G, \mathsf{T}\phi * f]=\sum_i A[a, H_i, \mathsf{T}\phi] \cdot A[b_i, G, \mathsf{T}f],$$

where b_i is an arbitrary point of H_i.

Remarks. (i) By the hypothesis there exist compact sets K_1 and K_2 such that $f(\bar{G}) \subset K_1$ and $\phi(\bar{H}) \subset K_2$. Since $\phi * f(\bar{G}) \subset f(\bar{G}) + \phi(\bar{H}) \subset K_1 + K_2$, then $\phi * f$ is also completely continuous on G.

(ii) The set of all roots of $\mathsf{T}\phi(x) = a$ is compact (a closed subset of $a - K_2$). Hence there are only a finite number of H_i such that $A[a, H_i, \mathsf{T}\phi] \neq 0$; thus the summation $\sum_i$ is to be taken only over those i such that $A[a, H_i, \mathsf{T}\phi] \neq 0$.

(iii) Since any two points b' and b'' in the same H_i can be joined by a polygonal line without touching $\mathsf{T}f(\bar{G} - G)$, it follows that $A[x, G, \mathsf{T}f]$ is constant in each H_i.

Proof of Theorem 9. Let D_k be the set of points $p \varepsilon H$ such that $A[p, G, \mathsf{T}f] = k$. Then we have

$$(1) \qquad \bar{D}_k - D_k \subset (\bar{H} - H) \cup \mathsf{T}f(\bar{G} - G),$$

and

$$\sum_i A[a, H_i, \mathsf{T}\phi] \cdot A[b_i, G, \mathsf{T}f] = \sum_k A[a, D_k, \mathsf{T}\phi] \cdot k.$$

Thus we have to prove

$$(*) \qquad A[a, G, \mathsf{T}\phi * f] = \sum_k A[a, D_k, \mathsf{T}\phi] \cdot k.$$

By the hypothesis and Theorem 1 there exists a $V \varepsilon \mathfrak{U}$ such that

$$(2) \qquad V(a) \cap (\mathsf{T}\phi(\bar{H} - H) \cup \mathsf{T}\phi * f(\bar{G} - G)) = 0.$$

From (1) and (2),

$$(3) \qquad V(a) \cap \mathsf{T}\phi(\bar{D}_k - D_k) = 0.$$

By Theorem 2 there exists a finite-dimensional linear manifold E^n and a continuous transformation S_2 of K_2 (Remark (i)) into E^n such that $S_2(x) - x \varepsilon V$ for $x \varepsilon K_2$. Thus

$$(4) \qquad \mathsf{T}S_2\phi(x) - \mathsf{T}\phi(x) \varepsilon V \text{ for } x \varepsilon \bar{H}.$$

Then by Corollary 3 we obtain from (3) and (4)

$$(5) \qquad A[a, D_k, \mathsf{T}\phi] = A[a, D_k, \mathsf{T}S_2\phi].$$

Now let X be the set of all roots of $\mathsf{T}S_2\phi(x) = a$. Then X is compact since $X \subset a - S_2(K_2)$. From (2) and (4) it follows that $a \bar{\varepsilon} \mathsf{T}S_2\phi\mathsf{T}f(\bar{G} - G)$.

Thus $X \cap \mathsf{T}f(\bar{G}-G)=0$. Then because X is compact and $\mathsf{T}f(\bar{G}-G)$ is closed, there exists a $U \varepsilon \mathfrak{U}$ such that

$$U(X) \cap \mathsf{T}f(\bar{G}-G)=0. \tag{6}$$

By Theorem 2 there exists finite-dimensional linear manifold E^m and a continuous transformation S_1 of K_1 into E^m such that $a \varepsilon E^m$ and $S_1(x)-x \varepsilon U$ for $x \varepsilon K_1$. Hence

$$\mathsf{T}S_1f(x)-\mathsf{T}f(x) \varepsilon U \text{ for } x \varepsilon \bar{G}. \tag{7}$$

Then by (6) and Corollary 3,

$$A[p, G, \mathsf{T}f]=A[p, G, \mathsf{T}S_1f] \text{ for } p \varepsilon X. \tag{8}$$

Now let D'_k be the set of points $p \varepsilon H$ such that $A[p, G, \mathsf{T}S_1f]=k$. Then by (8) the set $X \cap D_k$ of all roots of $\mathsf{T}S_2\phi(x)=a$ in D_k coincides with $X \cap D'_k$, that of all roots in D'_k. Thus by Corollary 1,

$$A[a, D_k, \mathsf{T}S_2\phi]=A[a, D'_k, \mathsf{T}S_2\phi] \quad (=A[a, D_k \cap D'_k, \mathsf{T}S_2\phi]).$$

Then by (5)

$$A[a, D_k, \mathsf{T}\phi]=A[a, D'_k, \mathsf{T}S_2\phi]. \tag{9}$$

On the other hand, from (2), (4), and $\mathsf{T}f(\bar{G}) \subset H$, we get by Corollary 3,

$$A[a, G, \mathsf{T}\phi * f]=A[a, G, \mathsf{T}(S_2\phi) * f]. \tag{10}$$

Now putting $(1-t)f+tS_1f=f_t$, we have $\mathsf{T}S_2\phi\mathsf{T}f_t=\mathsf{T}(S_2\phi)*f_t$, where $(S_2\phi)*f_t=(1-t)f+tS_1f+S_2\phi(\mathsf{T}f_t)$. Thus

$$(S_2\phi)*f_t(\bar{G}) \subset \hat{K}_1+\widehat{S_1(K_1)}+S_2(K_2)=K_3, \tag{11}$$

where K_3 is a compact set by Lemma 1. From (7) we get $\mathsf{T}f_t(x)-\mathsf{T}f(x) =t(S_1f(x)-f(x)) \varepsilon U$ for $0 \leqq t \leqq 1$. Thus by (6), $p \bar{\varepsilon} \mathsf{T}f_t(\bar{G}-G)$ for $p \varepsilon X$, $0 \leqq t \leqq 1$. Hence $a \bar{\varepsilon} \mathsf{T}(S_2\phi)*f_t(\bar{G}-G)$ for $0 \leqq t \leqq 1$. Thus by (11) and Theorem 7, $A[a, G, \mathsf{T}(S_2\phi)*f_t]$ is constant for $0 \leqq t \leqq 1$; hence $A[a, G, \mathsf{T}(S_2\phi)*f]=A[a, G, \mathsf{T}(S_2\phi)*(S_1f)]$, and by (10),

$$A[a, G, \mathsf{T}\phi * f]=A[a, G, \mathsf{T}(S_2\phi)*(S_1f)]. \tag{12}$$

Now let E^l be a finite-dimensional linear manifold containing both E^m and E^n. Then $S_2\phi(\bar{G}) \subset E^l$, $(S_2\phi)*(S_1f)(\bar{G}) \subset E^l$, and by the definition of the degree of mapping in E

$$\begin{cases} A[a, D'_k, \mathsf{T}S_2\phi]=A^l[a, D'_k{}^l, \mathsf{T}S_2\phi], \\ A[a, G, \mathsf{T}(S_2\phi)*(S_1f)]=A^l[a, G^l, \mathsf{T}(S_2\phi)*(S_1f)]. \end{cases} \tag{13}$$

But, since D'^l_k is just the set of points $p \varepsilon H$ such that $A^l[p, G^l, \mathsf{T}S_1 f] = k$, we obtain by Theorem 8

(14) $$A^l[a, G^l, \mathsf{T}(S_2\phi) * (S_1 f)] = \sum_k A^l[a, D'^l_k, \mathsf{T}S_2\phi] \cdot k,$$

using considerations similar to those in the first part of this proof Then from (9), (12), (13) and (14) follows (*).

6. Invariance of domain in complete metric E.

THEOREM 10. *Let E be complete metric, G an open set in E, and f a completely continuous transformation of $\bar{G}$ into E. If $\mathsf{T}f$ affords a one-one correspondence between $\bar{G}$ and $\mathsf{T}f(\bar{G})$, then $\mathsf{T}f(G)$ is an open set in E, and $A[b, G, \mathsf{T}f] = \pm 1$ for any $b \varepsilon \mathsf{T}f(G)$.*

To prove this theorem we shall first give two lemmas.

LEMMA 2. *If K is a compact set in a complete metric E, then the smallest closed convex set $\hat{K}$ containing K is also compact.*

Proof. Using the convexity of $\mathfrak{U}$ one can easily prove that $\hat{K}$ is totally bounded [10]; hence the closed set $\hat{K}$ is compact (in complete metric E).

LEMMA 3.[11] *If $f(x)$ is continuous on a closed set M in a separable complete metric E, and $f(M)$ is contained in a compact convex set K in E, then $f(x)$ can be extended to a continuous function on the whole of E in such a way that $f(E) \subset K$.*

Proof. Let $\{a_n\}$ be a denumerable set dense on M. We put

$$\rho(p, x) = \operatorname{Max}\{(2 - \operatorname{dist}(p, x)/\operatorname{dist}(p, M)), 0\} \text{ for } p \bar{\varepsilon} M;$$

then $\sum_{n=1}^{\infty} 2^{-n}\rho(p, a_n) f(a_n)$ converges uniformly on $E - M$,[12] and we define $f^*(p)$ by

$$f^*(p) = (\sum_{n=1}^{\infty} 2^{-n}\rho(p, a_n))^{-1} \sum_{n=1}^{\infty} 2^{-n}\rho(p, a_n) f(a_n) \text{ for } p \bar{\varepsilon} M,$$
$$= f(p) \text{ for } p \varepsilon M.$$

Then $f^*(p)$ is an extension of $f(p)$ as desired.

---

[10] For any $U \varepsilon \mathfrak{U}$ there exists a finite number of $p_i \varepsilon \hat{K}$ $(i = 1, \cdots, k)$ such that $\hat{K} \subset \bigcup_{i=1}^{k} U(p_i)$.

[11] This lemma I owe to Professor S. Kakutani.

[12] Because $\sum_{n=i}^{j} 2^{-n}\rho(p, a_n) f(a_n) \varepsilon 2^{-i+1}K$ (we assume $0 \varepsilon K$); hence for sufficiently large N, $\sum_{n=i}^{j} 2^{-n}\rho(p, a_n) f(a_n) \varepsilon U$ ($U \varepsilon \mathfrak{U}$ is arbitrarily given) if $i > N$, since K is compact.

Proof of Theorem 10. Let $b = \mathsf{T}f(a)$ be any point of $\mathsf{T}f(G)$, and let $f(\bar{G})$ be contained in a compact set K. By Lemma 2 we can assume that K is convex. Let E_0 be the smallest closed linear manifold containing K and a. Then E_0 is a separable complete metric subspace of E. Put $G_0 = G \cap E_0$, then $\bar{G}_0$ is transformed by $\mathsf{T}f$ into E_0. Then we have, since $b \,\bar{\varepsilon}\, \mathsf{T}f(\bar{G} - G)$, by the manner of definition of the degree of mapping,

$$(1) \qquad A[b, G, \mathsf{T}f] = A[b, G_0, \mathsf{T}f] \text{ in } E_0.$$

Now let us confine ourselves to E_0. Put $\mathsf{T}f(\bar{G}_0) = M$; then M is closed by Theorem 1. The inverse mapping $(\mathsf{T}f)^{-1}$ of M onto $\bar{G}_0$ is also continuous, because every closed subset of $\bar{G}_0$ corresponds to a closed subset of M by $(\mathsf{T}f)^{-1}$ (Theorem 1). Now put $x' = \mathsf{T}f(x)$, $(\mathsf{T}f)^{-1}(x') - x' = \phi_0(x')$; then $\phi_0(x') = -f(x)$, and hence $\phi_0(M) = -f(\bar{G}_0) \subset -K$. Therefore $\phi_0(x)$ is completely continuous on M, and $\mathsf{T}\phi_0 = (\mathsf{T}f)^{-1}$. Since M is closed in a separable complete metric E_0, and $\phi_0(M) \subset -K$ (a compact convex set in E_0), then by Lemma 3, $\phi_0(x)$ can be extended to a continuous $\phi(x)$ on the whole of E_0 in such a way that $\phi(E_0) \subset -K$.

Let H_i be the components of the open set $E_0 - \mathsf{T}f(\bar{G}_0 - G_0)$. Since $\mathsf{T}\phi^* f(x) = \mathsf{T}\phi_0 \mathsf{T}f(x) = x$ for $x \,\varepsilon\, \bar{G}_0$, then $a \,\bar{\varepsilon}\, \mathsf{T}\phi^* f(\bar{G}_0 - G_0)$. Thus by Theorem 9 we obtain, since $\bar{E}_0 - E_0 = 0$ and $A[a, G_0, \mathsf{T}\phi^* f] = 1$,

$$(2) \qquad 1 = \sum_i A[a, H_i, \mathsf{T}\phi] \cdot A[b_i, G_0, \mathsf{T}f],$$

where b_i is an arbitrary point of H_i. Consequently there exists an $i = k$ such that

$$(3) \qquad A[a, H_k, \mathsf{T}\phi] \cdot A[b_k, G_0, \mathsf{T}f] \neq 0.$$

Then by Theorem 5 there exists a $b_k \,\varepsilon\, H_k$ such that $a = \mathsf{T}\phi(b_k)$ and $b_k \,\varepsilon\, \mathsf{T}f(G_0) \subset M$. Thus $a = \mathsf{T}\phi_0(b_k) = (\mathsf{T}f)^{-1}(b_k)$; hence $b_k = \mathsf{T}f(a) = b$. Since there is only one H_k such that $b \,\varepsilon\, H_k$, we can assume $k = 1$ and $A[a, H_i, \mathsf{T}\phi] \cdot A[b_i, G_0, \mathsf{T}f] = 0$ for all $i > 1$. Thus from (2) follows

$$A[a, H_1, \mathsf{T}\phi] \cdot A[b, G_0, \mathsf{T}f] = 1.$$

Since the factors of the left side of this equation must be integers we get $A[b, G_0, \mathsf{T}f] = \pm 1$. Therefore from (1), returning to E,

$$A[b, G, \mathsf{T}f] = \pm 1.$$

Since $b \,\bar{\varepsilon}\, \mathsf{T}f(\bar{G} - G)$, there exists a $U \,\varepsilon\, \mathfrak{U}$ such that $U(b) \cap \mathsf{T}f(\bar{G} - G) = 0$.

Then by Corollary 2, $A[b', G, \mathsf{T}f] = \pm 1$ for any $b' \varepsilon U(b)$. Hence by Theorem 5, $U(b) \subset \mathsf{T}f(G)$. Thus the proof is complete.

OSAKA UNIVERSITY,
OSAKA, JAPAN.

REFERENCES.

[1] J. Leray and J. Schauder, "Topologie et équations fonctionelles," *Annales Scientifiques de l'École Normale Supérieure*, vol. 51 (1934), pp. 45-78.

[2] ———, "Topologie des espaces abstraits de M. Banach," *Comptes Rendus de l'Académie des Sciences* (Paris), vol. 200 (1935), pp. 1082-1084.

[3] M. Nagumo, "A theory of the degree of mapping based on infinitesimal analysis," *American Journal of Mathematics*, vol. 73 (1951), pp. 485–496.

[4] J. von Neumann, "On complete topological spaces," *Transactions of the American Mathematical Society*, vol. 37 (1935), pp. 1-20.

[5] E. Rothe, "The theory of topological order in some linear topological spaces," *Iowa State College Journal of Science*, vol. 13 (1939), pp. 373-390.

[6] J. Schauder, "Invarianz des Gebietes in Funktionalräumen," *Studia Mathematica*, vol. 1 (1929), 123-139.

[7] ———, "Über den Zusammenhang zwischen der Eindeutigkeit und Lösbarkeit partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Mathematische Annalen*, vol. 106 (1932), pp. 661-721.

[8] A. Tychonoff, 'Ein Fixpunktsatz," *Mathematische Annalen*, vol. 111 (1935), pp. 767-776.

[9] J. Wehausen, "Transformations in linear topological spaces," *Duke Mathematical Journal*, vol. 4 (1938), pp. 157-169.

THE TOPOLOGY OF THE LEVEL CURVES OF HARMONIC FUNCTIONS WITH CRITICAL POINTS.* [1]

By William M. Boothby.

Introduction. In a previous paper,[2] of which this is a continuation, topological properties of curve families which filled the Euclidean plane π, or a simply connected domain in π, were investigated. The families were assumed regular (i. e. locally homeomorphic to parallel lines) except at a possibly infinite collection of isolated singularities at each of which the family had the structure of a multiple saddle point; such families were called *branched regular curve families.* Further investigation of these families, in particular their relation to harmonic functions, is the aim of this paper. In what follows the definitions and theorems in [I] will be assumed, and the same notation will be used. In particular F, G will denote branched regular curve families filling the plane π, B will denote the set of singular points, R the domain $\pi - B$ in which F is regular, and so on. The Euclidean plane will be taken as a model for all simply connected domains.

The principal result of [I] was to prove that any branched regular curve family F filling π can be given as the family of level curves of a function $f(p)$ which is continuous on all of π and has no relative extrema. This generalizes a portion of [II] in which the same theorem is proved for a curve family without singularities in π. In this paper there are two main results: the first, proved in Section **1**, is that F is actually homeomorphic to the level curves of a harmonic function; the second, proved in Section **2**, asserts the existence of a decomposition of F into a countable collection of subfamilies of curves, each of which has the structure of the parallel lines $y =$ constant of the upper half-plane. Such subfamilies will be called half-parallel, and this decomposition has consequences for the study of harmonic functions and analytic functions which will be mentioned below. These two results generalize

* Received January 30, 1950.

[1] The material in this paper and the preceding paper was taken from the author's Ph. D. thesis at the University of Michigan. The author wishes to express his gratitude to Professor Wilfred Kaplan for his guidance in this research and his advice in the preparation of this paper.

[2] The Topology of Regular Curve Families with Multiple Saddle Points (pp. 405-438 of this volume). Theorems or section numbers preceded by I refer to this paper.

those of Kaplan [III] to curve families with singularities of the saddle point type.

The methods of proof used in [I] will also be of value here, i. e., it was first noted that the family F decomposes into single curves extending to infinity in each direction and collections of curves ending at branch points, where each collection forms a tree. Then it was shown that by removing enough curves we are left with a simply connected domain R^* in which the remaining curves F^* of F form a regular family. This allows us to use the theorems of Kaplan. By this method we are able to show in **1. 2** that for any branched regular curve family F, there exists a complementary family G, i. e. there is a branched regular curve family G with the same singularities as F and such that no pair of curves of F and G intersect more than once. For, from Kaplan [IV] it follows that for the family F^* filling R^* there is such a family G^*. Thus it remains only to modify G^* (along the curves removed from π to give us R^*) in such a fashion that the modified family $\tilde{G}^*$ becomes complementary to F when we replace the curves again. The details of this procedure are elaborated in **1. 2**.

Now given F, G complementary, by [I] there exist two functions f, g defined on π which have F, G as level curves. Using f and g, we may define a map $T: \pi \rightarrow uv$-plane by $T(p) = [f(p), g(p)]$ and since f, g have no extrema and have regular curve families as level curves it may easily be shown that T is light and interior. It follows from well known theorems that there is a homeomorphism $h: D \rightarrow \pi$, D a simply connected domain in the xy-plane, such that $w(x, y) = T[h(x, y)]$ is harmonic; but h maps the level curves of w onto F. This concludes in outline the proof of the first principal theorem. Since the converse is well known, this theorem give a characterization by *local* topological properties of the level curves of a function harmonic in a simply connected domain. We note an important but immediate corollary: F is homeomorphic to the family of solutions of a system of differential equations: $dy/dt = p(x, y)$, $dx/dt = q(x, y)$. Section **1. 3** gives us in detail the proof outlined above.

In Section 2 the decomposition theorem mentioned is proved, again by a reduction to theorems of Kaplan for families without singularities in a simply connected domain, in this case, as above, applied to the family F^* filling, and regular everywhere, in the domain R^* obtained by removing curves from F. The importance of this theorem lies in its possible applications to functions analytic in a simply connected domain, as follows. For example, let $u(x, y)$ be the real part of an entire function. The family F of its level curves will then be a branched regular curve family. Now for F by the

decomposition theorem there exists a countable set A and a decomposition of F into sets D_α, $\alpha \varepsilon A$, which are simply connected, overlap at most on their boundaries (and then along curves of F) and in each of which our entire function is 1-1. Thus the sets D_a furnish a decomposition of the Riemann surface of the inverse function into sheets. This is again a generalization of Kaplan [III] where these results were obtained for analytic functions, with non-vanishing derivative, defined in a simply connected domain.

1. The Branched Regular Curve Family as the Level Curves of a Harmonic Function.

1.1. Preliminary properties and definitions. In this paper a slightly more general definition of cross-section will be needed than in [I], as follows: a curve in R is a cross-section if every arc on it is a cross-section. This removes the restriction that a cross-section be an arc, i. e. it may extend to infinity in one or both directions or even be a bounded open or half-open curve.

Whitney [VI] has shown that in an orientable regular curve family filling a region S there is a function $f(p,t)$ with the properties: for each p in S and any t in $-\infty < t < \infty$, there is a unique point $q = f(p,t)$ lying on the curve C through p; $f(p,t)$ is continuous in both variables; $f(p,0) = p$ and as t increases (decreases) $f(p,t)$ moves continuously in the positive (negative) direction on C. Just as in [II] we have as an immediate corollary to this theorem the following:

THEOREM 1.1-1. *Let γ be a cross-section in F and $S(\gamma)$ the set of curves of F crossing γ. Then $S(\gamma)$ forms an open, simply connected set, F is regular in this set and, in fact there is a homeomorphism of $S(\gamma)$ onto* (i) *a strip* $0 \leqq y \leqq 1$, (ii) *a half-plane $y \geqq 0$ or* (iii) *the xy-plane, carrying the curves of F in $S(\gamma)$ onto the lines $y =$ constant, the case depending on whether* (i) *γ is an arc,* (ii) *γ is a half-open curve, or* (iii) *γ is an open curve, respectively.*

Proof. We will prove only case (ii), the proofs of the other cases being similar. It is a consequence of Theorem I 1.2-2 that $S(\gamma)$ is open and F is regular in $S(\gamma)$. From what follows it is clear that the domain filled by $S(\gamma)$ is simply connected, since it is homeomorphic to a half-plane. Let γ be parametrized by τ, i. e. $\phi(\tau)$ maps $0 \leqq \tau < \infty$ homeomorphically onto γ. Now $S(\gamma) - \gamma$ consists of two disjoint domains A and B and each curve, since it crosses γ exactly once, has an arc in A and an arc in B. Hence the curve family $F[S(\gamma)]$ is orientable, since we may assign to each C in this set the direction as positive along which we pass from A into B. Then by the

above mentioned theorem of Whitney we have a function $f(p, t)$ defined on S. Let $p = \phi(\tau)$, then the function $f(\phi(\tau), t)$ at once gives us, if we set $y = \tau$, $x = t$, a homeomorphism of $S(\gamma)$ onto the upper half plane with $x = 0$ the image of γ. This is essentially the same as the situation in [II, p. 174, Theorem 30].

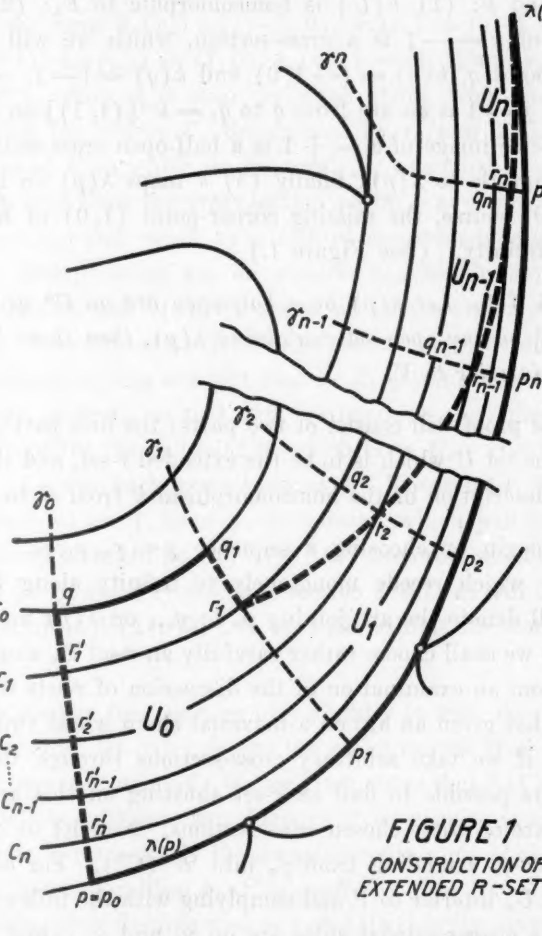


FIGURE 1
CONSTRUCTION OF EXTENDED R-SET

In [I] a useful "half-neighborhood" called an r-set was defined for any arc pq lying on a "maximal curve" C^*. The set was, briefly, a closed set in $\mathcal{D}^*(C^*)$ which abutted on pq and in which the curve family had the structure of the family F_1 of parallel lines, $y = k$, in a closed rectangle $R_1 = \{(x, y) \mid |x| \leq 1, 0 \leq y \leq 1\}$, the only singular points in this set being those on pq itself. Below we shall need a similar half-neighborhood for a half-open curve $\lambda(p)$ extending from a point p to infinity along a

maximal curve C^*. Of course, the most important case of such a situation is the "cut," $\lambda(b)$, from a branch point b. An *extended r-set*, $U(\lambda(p))$, of $\lambda(p)$ on C^* will then be a closed set contained in $\mathfrak{D}^*(C^*)$ together with a definite homeomorphism k of this closed set onto $\tilde{R}_1 = R_1 - \{(1,0)\}$, i. e. R_1 without its lower, left-hand corner point, where the following conditions are satisfied by U and k: (1) $F[U]$ is homeomorphic to F_1; (2) the inverse image under k of $x = -1$ is a cross-section, which we will denote by γ, joining p to a point q, $k(p) = (-1, 0)$ and $k(q) = (-1, -1)$; (3) the inverse image of $y = 1$ is an arc from q to $q_1 = k^{-1}[(1,1)]$ on a curve of F; and (4) the inverse image of $x = +1$ is a half-open cross-section Γ from q_1 to infinity, asymptotic to $\lambda(p)$; finally (5) k maps $\lambda(p)$ on the portion of $y = 0$ in $\tilde{R}_1$. Of course, the missing corner-point $(1, 0)$ of $\tilde{R}_1$ corresponds to the point at infinity. (See Figure 1.)

THEOREM 1.1-2. *Let $\lambda(p)$ be a half-open arc on C^* as defined above and let $V[\lambda(p)]$ be any open set containing $\lambda(p)$, then there is an extended r-set, $U(\lambda(p))$ interior to V.*

Proof. The proof will consist of two parts: the first part (A) being the description of the set U which is to be the extended r-set, and the second part (B) being the description of the homeomorphism k from U to $\tilde{R}_1$. (Fig. 1.)

(A) We begin by choosing a sequence $p = p_0, p_1, p_2, \cdots$ of regular points on $\lambda(p)$ which recede monotonely to infinity along that half-open curve. We shall denote the arc joining p_n to p_{n+1} on $\lambda(p)$ merely by p_np_{n+1}, and as first step we shall choose rather carefully an r-set U_n abutting on p_np_{n+1} for each n. From an examination of the discussion of r-sets in [I], it is not difficult to see that given an arc on a maximal chain whose two endpoints are regular points, if we take arbitrary cross-sections through these endpoints, then it is always possible to find an r-set abutting on this arc whose cross-sectional sides are on these chosen cross-sections. Now let us choose for each n a cross-section γ_n extending from p_n into $\mathfrak{D}^*(C^*)$. For each n we shall choose the r-set U_n interior to V and complying with the following conditions: first, so that its cross-sectional sides are on γ_n and γ_{n+1}, and so that it lies within an ϵ_n-neighborhood of p_np_{n+1}, $\epsilon_n \to 0$. And second, having chosen U_{n-1}, and denoting by q_np_n the arc on γ_n which forms that one of the cross-sectional sides of U_{n-1} on γ_n, we choose U_n in such a manner that its cross-sectional side r_np_n along γ_n is contained in q_np_n, i. e. r_n lies between q_n and p_n on γ_n. Then $U_{n-1} \cap U_n = r_np_n$ by Theorem I 3.2-2, and C_{r_n} (denoted below by C_n) intersects γ_0 at some point r'_n. Of course, each U_n is in $\mathfrak{D}^*(C^*)$ since each γ_n is. Thirdly, we require the U_n so chosen that r_n is contained in a

δ_n-neighborhood of $p = p_0$ for a definite sequence $\delta_n \to 0$. The properties of r-sets as developed in [I] make it obvious that these conditions on U_n can be complied with. (Note: below we denote γ_0 simply by γ.)

Let k_n denote the homeomorphism of U_n onto R_1; we shall always assume k_n so chosen that as we move from p towards infinity on $\lambda(p)$ the image point moves from left to right along the x-axis. Then if we consider the image of U_n in R_1, we have $k_n(p_n) = (-1, 0)$, $k_n(r_n) = (-1, -1)$, $k_n(p_n r_n) =$ (line $x = -1$); $k_n(p_{n+1}) = (1, 0)$, $k_n(q_{n+1}) = (1, 1)$ and $k_n(r_{n+1}) = (1, a)$ where $0 < a < 1$. Now it is clear, as noted above, from the description of the sets U_n, that the curves C_n determined by r_n and q_{n+1}, $n = 1, 2, \cdots$, i. e. those curves of F on which lie the arcs $r_n q_{n+1}$, carried by k_n onto the line $y = +1$ in R_1, will cross the cross-section pq on γ at points r'_n which form a monotone sequence and which by our third requirement approach p. Also the points qq_1 determine an arc on a curve C_0, which maps under k_0 on $y = +1$. It follows that $S(\gamma)$, the set of all curves crossing γ, will contain all the sets U_n. Now for each n let a cross-section $r_n r_{n+1}$ be determined in U_n as the inverse image of the straight line in R_1 joining $k_n(r_n)$ and $k_n(r_{n+1})$. If we direct each curve C_n so that $\mathfrak{D}^\#(C_n)$ contains $\lambda(p)$ then $r_{n-1}r_n$ will lie in $\mathfrak{D}^*(C_n)$, except for the endpoint r_n which is on C_n. Hence the arcs $q_1 r_1$, $q_1 r_1 r_2$, $q_1 r_1 r_2 r_3, \cdots$ are each cross-sections by Theorem I 3.5-3, and they approach as limit an arc Γ from q_γ to infinity, which will be then a cross-section. The set U bounded by (1) $\lambda(p)$, (2) the cross-sectional arc pq on γ, (3) the arc qq_1 on C_0 and (4) the cross-section Γ will, as will be shown below, be an r-set interior to $V(\lambda(p))$, i. e. we shall exhibit the homeomorphism k of the definition.

(B) Now change the meaning of γ slightly to let γ denote only the arc pq on γ and, as above, let $S(\gamma)$ denote the set of curves of F crossing γ, and $S(\Gamma)$ the set of curves crossing Γ. We see from above that $S(\Gamma) \cup C^* = S(\gamma)$. Each of these sets is split into two domains if we remove γ, and we shall let $S'(\Gamma)$, $S'(\gamma)$ denote respectively the domains containing $\lambda(p)$. By Theorem 1. 1-1 there is a homeomorphism $k_1 : S'(\gamma) \to R_1''$, $R_1'' = \{(x, y) \mid -1 \leqq x < \infty, 0 \leqq y \leqq 1\}$ and $k_1(\gamma)$ is the line $x = -1$, $k_1(\lambda(p))$ is the x-axis for $x \geqq -1$. The curve C_0 on which q, q_1 lie will map onto the line $y = 1$ and Γ'', the image of Γ, will be an arc given by $x = \phi_2(y)$, $0 \leqq y < 1$, $k_1(q_1) = (\phi_2(1), 1)$ and $\lim_{y \to 0} \phi_2(y) = \infty$.

Now we let $R_1' = \{(x, y) \mid -1 \leqq x < 1, 0 \leqq y \leqq 1\}$, i. e. a rectangle with the right side missing. We shrink $S'(\gamma)$ into R_1' along the lines $y =$ constant by the homeomorphism $k_2 : S'(\gamma) \to R_1'$ where $k_2 : (x, y) \to (x, y)$

3

for $-1 \leqq x \leqq 0$ and $k_2: (x, y) \to (x/(x+1), y)$ for $0 \leqq x < +\infty$. Let $x = \phi_1(y)$ be the function whose graph is Γ', the image of Γ'' under k_2. Γ' will be a half-open arc from the point $(\phi_1(1), 1)$ on the line $y = 1$ to the point $(1, 0)$ as limit point, i. e. as $y \to 0$ the point $(\phi_1(y), y)$ approaches $(1, 0)$ along Γ'. The side $x = -1$, the side $y = 0$, the segment from $(-1, 1)$ to $(\phi_1(1), 1)$ on $y = 1$, and finally Γ' will bound the image of U under k_2k_1. Denote this portion of R_1' by U'. Then we perform a final homeomorphism $k_3: U' \to \hat{R}_1$ where $k_3: (x, y) \to \{(2 + 2x)/(1 + \phi_1(y)) - 1, y\}$. The combined homeomorphism $k = k_3k_2k_1: U \to R_1$ is easily seen to be the desired homeomorphism as required in the definition of an extended r-set. Hence the proof of the theorem is complete. With these preliminaries completed we can prove the existence of the family G, complementary to F.

1. 2. Complementary curve families. Given a branched regular curve family F filling π, we shall call another such family, G, filling π *complementary* to F if (1) the singularities of G are exactly those of F and each is of the same type, i. e., a point b is an n-th order branch point of G if and only if it is an n-th order branch point of F; and (2) every curve of G is a cross-section of F. It follows at once from this definition and Theorem I 3. 2-4 that if G is complementary to F, then F is complementary to G. Hence we may speak of two complementary families, F and G, filling π. They will have a common set of singular points, B.

The major result of this section is to establish that every branched regular curve family F has a complementary family G. In [IV] it is shown that this is true when the set of singular points is empty, i. e. for a family F^* regular throughout a simply connected domain R^*, for we may by [IV] map F^* onto a family F' filling the xy-plane and defined by differential equations, $dx/dt = f(x, y)$; $dy/dt = g(x, y)$. The orthogonal trajectories define a family G' complementary to F' and the inverse image G^* of G' is then the desired complementary family to F^*. This result immediately gives us a family G^* complementary to F^* in $R^* = \pi - \tilde{\mathfrak{I}}$, $\tilde{\mathfrak{I}}$ being the cuts removed from π to make R simply connected [I]. The method we shall use to establish the existence of a family G complementary to F will be to consider first F^* and its complementary family G^*, both defined in R^* and then to modify G^* slightly near the boundary of R^*, i. e., near the cuts $\lambda(b)$, obtaining a family $\tilde{G}^*$ which becomes a family $\tilde{G}$ of the desired type when $\tilde{\mathfrak{I}}$, the set removed from π to give R^*, is replaced again. Theorem I 4. 1-3 tells us that we may cover $\tilde{\mathfrak{I}}$ with a collection $\{V[\lambda(b)]\}$ of disjoint open sets; we shall assume such a covering, to be fixed throughout what follows, and moreover,

assume that each $V \subset U_\epsilon[\lambda(b)]$ an ϵ-neighborhood of $\lambda(b)$ where $\epsilon > 0$ is fixed. Any modification in G^* will actually take place deep inside V, i. e., in an open set whose closure lies in V. We shall actually discuss the modification for one such V and, assuming similar modifications have taken place in each V, we will denote by $\tilde{G}^*$ the modified G^*. $\tilde{G}^*$ will then be shown to be such that when $\tilde{\mathfrak{J}}$ is replaced, $\tilde{G}^*$ becomes a set $\tilde{G}$ complementary to F.

Restricting ourselves then to a definite $\lambda(b)$, and its neighborhood $V[\lambda(b)]$, as model, we proceed to define for $\lambda(b)$ a certain possibly infinite collection of closed sets $W_0, W_1; \cdots$ all contained inside $V[\lambda(b)]$ and surrounding $\lambda(b)$. These are the sets in which G^* will be modified. W_0 is an extended r-set, and if the number of curves in $\lambda(b)$ is finite, and only in this case, there will be a last set W_N of this collection which is also an extended r-set. All the other sets W_i will be r-sets in the sense of [I], which we shall hereafter call merely r-sets. These sets will be chosen as follows: *First*, let $b_0 = b, b_1, b_2, \cdots$ be the branch points on $\lambda(b)$, numbered so as to recede monotonely to infinity, and let the curves in R^* of each $St(b_i)$ be numbered with two indices, the first being that of b_i, the second being given by a counterclockwise numbering of the $St(b_i)$ proceeding from the first curve which follows counterclockwise after a curve of $St(b_i)$ not on C^* ($C^* \supset \lambda(b)$) to the last curve of $St(b_i)$ not on C^*: $C_{01}, \cdots, C_{0n_1}; C_{11}, C_{12}, \cdots, C_{1n_2}; \cdots$ etc.; let C_{00} denote C. (See Figure 2.) Then all numbered curves are in R^*. *Second*, choose regular points s_{ij} on each C_{ij} and short cross-sections γ_{ij} through s_{ij}, the γ_{ij} being in each case an arc on a curve of G^* and both s_{ij} and γ_{ij} being chosen so as to lie in $V(\lambda)$. Now we choose our sets W_n as follows: $W_0 \subset V(\lambda)$ is an extended r-set bounded on one side by an arc $r_{00}s_{00}$ on γ_{00} and on one side, of course, by $(s_{00}b) \cup \lambda(b)$. Next, in the domain bounded by the maximal chain determined by the adjacent curves C_{00}, C_{01} we choose an r-set $W_1 \subset V(\lambda)$ of the arc $s_{00}s_{01}$ on these curves, with W_1 bounded by the arcs $s_{00}t_{00}$ on γ_{00} and $r_{01}s_{01}$ on γ_{01}. Similarly, we choose $W_2, \cdots, W_{n_1}$, each an r-set contained in $V(\lambda)$ and bounded by arcs on two of the γ_{0i}'s. It may be that b_0 is the only branch point of $\lambda(b)$, in which case the next set W_{n_1+1} is the last and must be an extended r-set, bounded on one side by an arc $s_{0n_1}t_{0n_1}$ on γ_{0n_1}. Otherwise, we choose for W_{n_1} an r-set of $s_{0n_1}b_0b_1s_{11}$, an arc on the adjacent chain C_{1n_1}, C', C_{11}; C' being the curve of $\lambda(b)$ with endpoints b_0, b_1. The r-set W_{n_1} is so chosen that its cross-sectional ends are arcs $s_{0n_1}t_{0n_1}$ and $r_{11}s_{11}$ on γ_{0n_1} and γ_{11} respectively, and that it lies in $V(\lambda)$. This process is continued until we have chosen r-sets (or extended r-sets) on both sides of every curve of $St(b_i)$ for all b_i and hence, in particular, on both sides of

each curve of $\lambda(b)$. Then $\lambda(b)$ will be contained in the interior of the set $W_\lambda = \bigcup_i W_i$. W_λ is bounded by an open arc Γ extending to infinity in each direction; and Γ consists in one case of *one* infinite cross-section of F^*, not in general a curve of G^*, plus an infinite number of arcs alternately on curves of F^* and on curves of G^* (the latter of the form $r_{ij}s_{ij}t_{ij} \subset \gamma_{ij}$); or else in

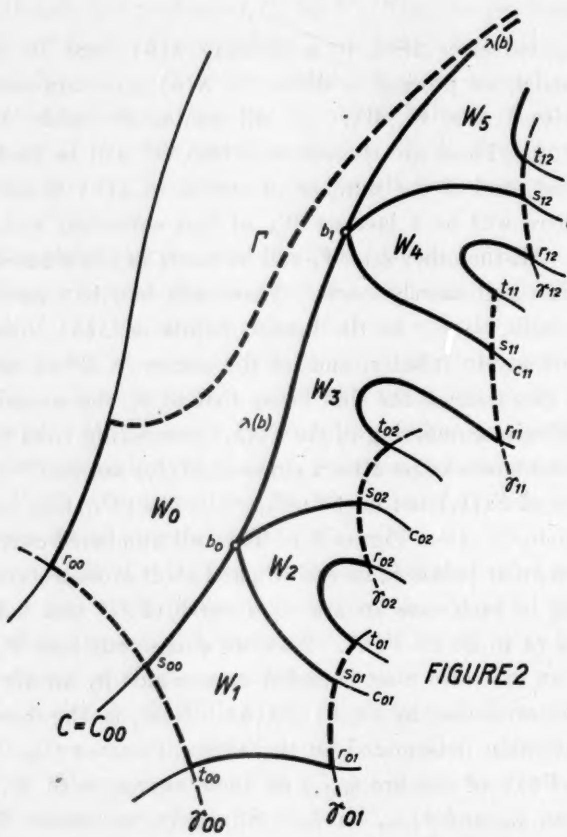


FIGURE 2

the other case, Γ consists of a finite number of such alternate arcs on F^* and G^* plus *two* half-open cross-sections of F^* extending to infinity. The first case occurs when there is one extended r-set and the number of sets W_i is infinite, the second when there are two extended r-sets and a finite collection of sets W_i comprising W_λ. Γ lies entirely inside $V(\lambda)$, and W_λ, which consists of Γ plus that one of its complementary domains inside $V(\lambda)$, is a closed set. The W_i's clearly intersect only on curves of F, namely on $\lambda(b)$ and on the arcs b_is_{ij} on each curve C_{ij} of $R^* \cap St(b_i)$ for b_i in $\lambda(b)$. We

denote by $\bar{\lambda}$ the set of all points which lie on the common boundary of two or more W_i's. A point of $\bar{\lambda}$ which is a regular point clearly lies on the intersection of just two such sets, whereas each branch point b_i lies on the intersection of $2m$, where m is the multiplicity of b_i. We denote by W_i^* the set $W_i - \bar{\lambda}$, and by W_λ^* the set $W_\lambda - \bar{\lambda}$, and finally by V_λ^* the set $V(\lambda) - \bar{\lambda}$. Then let $\bar{G}^* = G^*[V_\lambda^*]$ and $\bar{F}^* = F^*[V_\lambda^*]$, i. e. we remove from $V(\lambda)$ all points on two or more sets W_i.

Now each W_n has associated with it a homeomorphism k_n, of W_n onto R_1 or, if it is an extended r-set, onto $\tilde{R}_1$. In order that the modification of G^* to $\hat{G}^*$ which we are going to make will not destroy the relationship between G^* and F^* we will actually achieve it by a homeomorphism h of $\tilde{R}^*$ ($\tilde{R}^* = R^* - \bigcup_{\lambda \varepsilon \tilde{T}} \tilde{\lambda}$) onto itself, which is the identity outside of each set W, but which inside such a set carries each curve of $\bar{F}^*$ onto itself. The need for this modification arises from the fact that although $\bar{F}^*$, $\hat{G}^*$ are complementary in $\tilde{R}^*$, and hence in $V^*(\lambda) = V(\lambda) - \tilde{\lambda}$, they will not in general be complementary in $V(\lambda)$ along points of $\bar{\lambda}$. In fact replacing $\bar{\lambda}$ in V^* will not in general transform $\bar{G}^*$ into a regular curve family in $V(\lambda)$, since after all those points of $\bar{\lambda}$ on λ are boundary points of the region R^* filled by G^*, and curves of G^* may have common endpoints on the boundary of the domain of G^*, or no endpoints (i. e. may extend to ∞ which is also a boundary point of R^*). Our procedure is to cut the plane along each $\tilde{\lambda}$, cutting along curves of F, i. e. whenever we cut a curve of G^*, we cut across it: in particular in cutting along $\bar{\lambda} - \lambda$, since this is in R^*, we cut across curves of G^*. Then keeping the curves of F fixed (not pointwise) we move the "cut ends" of curves of G^*, with their individual points "sliding" along curves of F, into such positions that each regular point on $\bar{\lambda}(b)$ becomes the endpoint of exactly one curve of G^* from each side of $\tilde{\lambda}$, and the branch points of multiplicity m the endpoint of $2m$ curves, one from each sector. Then replacing $\bar{\lambda}$. $\hat{G}^*$ the modified G^* becomes a regular family at every regular point of F and is in fact complementary to F. We shall describe this operation piecewise, for each W_n^* and, in fact, at first as a homeomorphism on the image of W_n^* in R_1 (or $\tilde{R}_1$ as the case may be), (I) for r-sets, (II) for extended r-sets.

(I) We begin by defining a typical homeomorphism f_I in R_1 on the image under k_i of $\bar{F}^*[W_i^*]$, $\bar{G}^*[W_i^*]$, W_i an r-set. The image of W_i^* will be $R_1^* = R_1 - (x\text{-axis})$, and we will denote the images of the curve families as F_1^*, G_1^*, respectively. The former will, of course, be just the lines $y = a$, $0 < a \leqq 1$, and the latter will be a regular curve family filling R_1^*, complementary to F_1^*, and having among its curves the two lines $x = \pm 1$, images

of arcs γ_{ij}, which lie on curves of G^*. It will be seen that $G_1{}^*$ consists exactly of the curves whose inverse images cross C', the inverse image of $y = 1$ in $R_1{}^*$, for, if we consider any curve of $\bar{G}^*$ with a point inside W_i, it is clear that it must leave W_i in each direction, there being no boundary points of $\tilde{R}^*$ interior to W_i; and hence, it must either cross C' or have two endpoints on $\tilde{\lambda}(b)$. It could scarcely have both endpoints on $\tilde{\lambda}(b)$, however, without crossing some curve of F^* twice inside W_i, which is impossible since the curves of G^* are cross-sections of F^*. Moreover, no curve of G^* will cross C' more than once, since C' is a cross-section of G^*. Thus we may define a function f_I mapping $R_1{}^*$ onto itself as follows: Let $\bar{x} = f(x, y)$ be defined by $f(x, 1) \equiv x$ and $f(x, y) =$ constant on each curve of $G_1{}^*$, and let $\bar{y} = g(x, y)$ be defined by $g(x, y) \equiv y$. Then it follows from the above remarks and the work of Kaplan [II] and [III] that $f_I\colon (x, y) \to (\bar{x}, \bar{y})$ is a homeomorphism of $R_1{}^*$ onto itself which takes each curve of $F_1{}^*$ onto itself, and each curve of $G_1{}^*$ onto a line $x = b$, $-1 \leqq b \leqq 1$, the lines $x = \pm 1$ being held pointwise fixed, as is the line $y = 1$, i.e., all of the boundary of $R_1{}^*$ on which f_I is defined is held pointwise fixed. $h \mid W_i{}^*$ is then defined by $k_i{}^{-1} f_I k_i$, and if thus defined h maps $\bar{F}^*[W_i{}^*]$ onto itself, takes $\bar{G}^*[W_i{}^*]$ homeomorphically onto a new family $\tilde{G}^*[W_i{}^*]$ which is still complementary to F^* and which is identical to $\bar{G}^*$ on the boundary of $W_i{}^*$. Since k_i is actually a homeomorphism of all of W_i onto R_1, it will now map $F[W_i]$ and $\tilde{G}^*[W_i]$ so that the curves $F^*[W_i]$, $\tilde{G}^*[W_i]$ will map onto the lines $y = a$ and $x = b$, respectively. We re-denote k_i by $\tilde{k}_i$ to emphasize that it acts on $\tilde{G}^*$. Thus it is clear that every curve of $\tilde{G}^*[W_i]$ has exactly one endpoint, unique to it, on $\tilde{\lambda}$ and exactly one endpoint unique to it on the curve of F^* forming the opposite side of W_i. The regularity of $\tilde{G}^*$ which we have achieved at $\tilde{\lambda}$ is precisely what is needed. We assume a similar homeomorphism defined for every index i such that W_i is an r-set; then h will be defined on every set of W except the one (or two) extended r-set(s).

(II) Now let us suppose that we are dealing with an extended r-set say W_0, with its associated homeomorphism k_0 onto $\tilde{R}_1$. Again let $F_1{}^*$, $G_1{}^*$ denote the images of the respective families of W_0 in $\tilde{R}_1{}^* = \tilde{R}_1 - (x\text{-axis})$, $F_1{}^*$ being just the lines $y = a$; the line $x = -1$ in $\tilde{R}_1{}^*$, but not in general the line $x = +1$, being a curve of $G_1{}^*$. f_{II} will be given as the composition of four homeomorphisms of $\tilde{R}_1{}^*$ onto itself. Before we can describe f_1, the first of these, we must note that there is in W_0 at least one curve ψ of G^*, distinct from the arc $r_{00}s_{00}$ on γ_{00} (the inverse image of $x = -1$), whose image ψ_1 in $\tilde{R}_1$ joins a point $(x'', 0)$ to a point $(x', 1)$, where $-1 < x'', x' < 0$,

i. e., a curve of G^* joining one side of W_0 to the other, and intersecting each side at a regular point of R^*, in particular, on C_{00}, not on $\lambda(b)$ (see Figure 3). That such a curve exists follows from the fact that in the family G^*, regular in R^*, the arc $r_{00}s_{00}$ on a curve of G^* has an r-neighborhood U (by Theorem I 1. 2-2) with $\bar{U} \subset R^*$. The curves $C_{s_{00}}$ and $C_{r_{00}}$ (see Figure 2) have small arcs entirely in this neighborhod, since they are cross-sections of G^*, and each of these will be crossed by an infinite number of curves of G^* on each side of $s_{00}t_{00}$, one of which will serve our purpose; namely, one crossing for each of these arcs that part which is the inverse image respectively of the

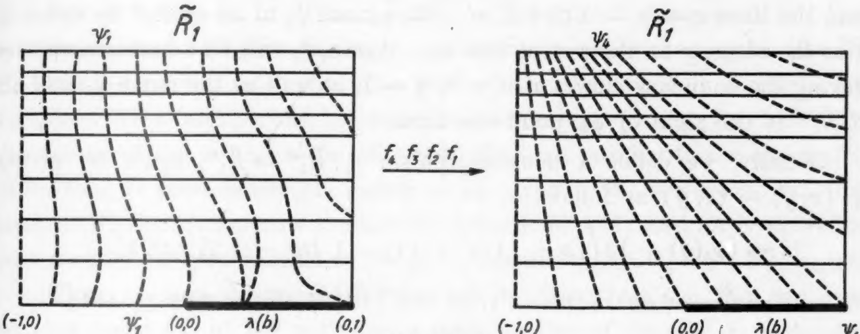


FIGURE 3

segments $(-1, 1)$ to $(0, 1)$ and $(-1, 0)$ to $(-\epsilon, 0)$, $1 > \epsilon > 0$. ψ_1 will be given by a continuous function $x = \psi_1(y)$, $0 \leqq y \leqq 1$, and we shall use it to define $f_1: R_1^* \to R_1^*$ given by $f_1: (x, y) \to (\bar{x}, \bar{y})$ where

$$\bar{x} = \{[1 + \psi_2(y)]x - [\psi_1(y) - \psi_2(y)]\}/[1 + \psi_1(y)] \quad \text{for } -1 \leqq x \leqq \psi_1(y),$$

$$\bar{x} = \{[1 - \psi_2(y)]x - [\psi_1(y) - \psi_2(y)]\}/[1 - \psi_1(y)] \quad \text{for } \psi_1(y) \leqq x \leqq +1;$$

$$\bar{y} \equiv y$$

(where $\psi_2(y) = (x' - x'')y + x''$, this being the equation of the line joining $(x', 1)$ to $(x'', 0)$, the curve into which ψ_1 is mapped by f_1).

The next homeomorphism, $f_2: R_1^* \to R_1^*$ will carry ψ_2 into ψ_3, the line $x = x'$. f_2 is given by $f_2: (x, y) \to (\bar{x}, \bar{y})$ where:

$$\bar{x} = \{(1 + x')x + [x' - \psi_2(y)]\}/[1 + \psi_2(y)] \quad \text{for } -1 \leqq x \leqq \psi_2(y),$$

$$\bar{x} = \{(1 - x')x + [x' - \psi_2(y)]\}/[1 - \psi_2(y)] \quad \text{for } \psi_2(y) \leqq x \leqq 1;$$

$$\bar{y} \equiv y.$$

Each of these homeomorphisms holds the boundary curves $x = \pm 1$, $y = 1$ pointwise fixed. To describe f_3 we first denote by M that portion of $\tilde{R}_1^*$ which lies on or to the left of ψ_3, i. e., $M = \{(x, y) \mid -1 \leqq x \leqq x', 0 \leqq y \leqq 1\}$. M is bounded on each side by a line $x =$ constant, i. e. $x = -1$, and $x = x'$, each the image of a curve of G^* under the composition of the above maps, and M is bounded on top and bottom by images of curves of F^*. The image of F^* in M is the family of lines $y = a$. Hence by precisely the same argument as in the definition of f_I for the neighborhood of type I above, we may find a homeomorphism $f_3 : M \to M$ which holds the boundary of M pointwise fixed, takes each curve $y = a$ onto itself, and takes the image family of G^* onto the lines $x = b$, $-1 \leqq b \leqq x'$. We extend f_3 to all of R_1^* by defining it as the identity on the rest of this set. Again, f_3 will be a homeomorphism leaving the boundary curves $x = \pm 1$, $y = 1$, as well as the curve ψ_3 and all of $\tilde{R}_1^*$ to the right of ψ_3, pointwise fixed.

Finally, we define a homeomorphism $f_4 : \tilde{R}_1^* \to \tilde{R}_1^*$, again by giving $f_4 : (x, y) \to (\bar{x}, \bar{y})$ as follows:

$$\bar{x} = (\psi_4(y) + 1)((x + 1)/x' + 1)) - 1 \text{ for } -1 \leqq x \leqq x',$$

$$\bar{x} = (1 - \psi_4(y))((x - x')/(1 - x')) + \psi_4 \text{ for } x' \leqq x \leqq +1;$$

$$\bar{y} = y,$$

where ψ_4 denotes the line $x = \psi_4(y) = (x' - 1)y + 1$ joining $(x', 1)$ to $(1, 0)$, this being the image of ψ_3 under f_4. The image of M under f_4 will be denoted by M_1 and will be the trapezoid bounded by ψ_4, the x-axis, the line $x = -1$, and the segment from $(-1, 1)$ to $(x', 1)$ on the line $y = 1$. f_4 takes the lines $y = a$ onto themselves, and the lines $x = b$, $-1 \leqq b \leqq x'$ of M onto a family of nonintersecting straight lines joining the points of the top edge of M_1 to the bottom (as listed above). f_4 leaves the lines $x = \pm 1$ and $y = 1$ pointwise fixed.

Now we define $f_{II} : \tilde{R}_1^* \to \tilde{R}_1^*$ as the homeomorphism $f_4 f_3 f_2 f_1$, and we define $h \mid W_0^*$ as $k_0^{-1} f_{II} k_0$ (see Figure 3). Then $h \mid W_0^*$ is a homeomorphism of $W_0^* = W_0 - \bar{\lambda}$ onto itself which is pointwise fixed on the boundary of W_0^* in $\tilde{R}^*$, i. e., on $t_{00}s_{00}$, on $C_{t_{00}}$, and on the extended cross-section which bounds one side of W_0. h also takes the curves of $G^*[W_0^*]$ homeomorphically onto a family $\tilde{G}^*$, at the same time mapping each curve of F^* onto itself. Now, if as above for k_i, we re-denote k_0 by $\tilde{k}_0$, then we have a homeomorphism of all of W_0 onto $\tilde{R}_1$ which takes $\bar{\lambda}$ onto the x-axis between $(-1, 0)$ and $(1, 0)$, with b_0 mapping onto $(0, 0)$, and s_{00} onto $(-1, 0)$, and which moreover, takes the curves of F onto the lines $y = a$ and takes part of $\tilde{G}^*$ onto the

straight lines joining the top and bottom of M_1 as described above, the remainder of $\tilde{G}^*$ mapping onto a regular family filling the rest of R_1. The curve Ψ of $\tilde{G}^*$, image of ψ under $h \mid W_0{}^*$ divides W_0 into two domains, one of which maps onto M_1, the other onto $R_1 - M_1$. We shall denote the one which maps onto M_1, together with its boundary, by $\tilde{W}_0$, the boundary consisting of two curves of $\tilde{G}^*$, namely $r_{00}s_{00}$ and Ψ, together with $C_{r_{00}}$ and $s_{00}b_0 \cup \lambda(b)$ in F. It is obvious that $M_1 \subset \tilde{R}_1$ can be mapped onto $\tilde{R}_1$ by a homeomorphism g which holds $x = -1$ and $y = 0$ pointwise fixed, takes each line $y = a$ into itself, and finally moves the image curves of $\tilde{G}^*$ in M_1 onto the lines $x = b$, $-1 \leqq b \leqq 1$, keeping, of course, their lower endpoints fixed, thus taking the line Ψ onto $x = 1$. Then $g\tilde{k}_0 \colon \tilde{W}_0 \to \tilde{R}_1$ with F going onto the lines $y =$ constant and $\tilde{G}^*$ onto the lines $x =$ constant. $\tilde{W}_0$ is then again, like W_0, an extended r-set of $\lambda(b)$, but of a kind which is bounded by curves of two complementary families and has associated a homeomorphism $g\tilde{k}_0$ which maps the curves of the respective families onto the lines parallel to the axes in $\tilde{R}_1$. Hereafter, we shall denote $g\tilde{k}_0$ merely by $\tilde{k}_0$. Now if W_N is a second extended r-set in W, then it must be the last W_i defined for $\lambda(b)$ and on it we define, in a manner entirely parallel to the above discussion, f_{II}, $h \mid W_N{}^*$, $\tilde{W}_N$, $\tilde{k}_N$, etc.

Thus we have defined $h \mid W_i{}^*$ for all i. Now since the $W_i{}^*$ are overlapping closed sets of $V_\lambda{}^*$ with only a finite number of the sets W_i containing any given point, and since h is actually the identity along their overlapping boundaries as well as on Γ, the boundary of W_λ, we have defined a homeomorphism h of $W_\lambda{}^* = W_\lambda - \bar{\lambda}$ onto itself. Assume that h is similarly defined for a set $W_\lambda{}^* \subset V[\lambda(b)]$ for *every* cut $\lambda(b)$ contained in $\tilde{\mathfrak{T}}$, and define h as the identity outside the W_λ's. We remark that the collection of all the sets W_λ for $\lambda(b)$ in $\tilde{\mathfrak{T}}$, together with the set $\pi - \bigcup_{\lambda \epsilon \tilde{T}} W_\lambda$, is a collection of overlapping closed sets which has a locally finite character, i. e., every neighborhood of any point meets only a finite number of the closed sets. This is clear because the cuts, λ, recede to infinity, and each W_λ lies in an ϵ-neighborhood of the cut λ, $\varepsilon > 0$ being fixed. Then it follows that h is a homeomorphism of $\tilde{R}^*$ onto itself, where by $\tilde{R}^*$ we mean $R^* - [\bigcup_{\lambda \epsilon \tilde{T}} \bar{\lambda}(b)]$. h carries every curve of F^* onto itself homeomorphically, and every curve of $G^*[R^*]$ homeomorphically onto a family $\tilde{G}^*$ which is complementary to F^* in $\tilde{R}^*$ and which coincides with G^* except in the interior of the W_λ's.

It remains to prove that by adding the boundary points of $\tilde{R}^*$, i. e., $\bigcup_{\lambda \epsilon \tilde{T}} \bar{\lambda}$, the curves of $\tilde{G}^*$ become curves of a family $\tilde{G}$ complementary to F in π. To prove this we must first prove that $\tilde{G}$ is regular in $R = \pi - B$. Now if v is a point of $\tilde{R}^*$, this is clear, since $\tilde{G} = \tilde{G}^*$ (which is homeomorphic to G^*)

in some neighborhod of p. In fact, it is clear (from the method used in getting the homeomorphism f_I above) that there is an arbitrarily small r-neighborhod of p whose closure maps onto $R_0 = \{(x, y) \mid |x| \leqq 1, |y| \leqq 1\}$ so that the lines $x =$ constant are the images of the curves of $\tilde{G}$, those lines $y =$ constant are the image curves of F.

Now, however, suppose that p is a regular point on $\tilde{\lambda}(b)$. Then p will be on the common boundary of just two of the neighborhoods W_i, since p is not a branch point. Let W_n, W_m be the two neighborhoods. Then p is interior to $W_n \cup W_m$, and it follows from Theorem I 1. 2-3 that $\tilde{G}[W_n \cup W_m]$ is regular at p, since $\tilde{G}$ is regular in W_n and in W_m separately, as may be seen from the existence of the maps $\tilde{k}_n, \tilde{k}_m$ onto R_1 (or $\tilde{R}_1$ as the case may be) with $\tilde{G}$ mapping onto the lines $x =$ constant. Thus $\tilde{G}$ is regular at every point of R, so that the singularities of $\tilde{G}$ are contained in the set B of singularities of F, and are thus isolated. Now each branch point is in a cut, and hence will be $b_i \,\varepsilon\, \lambda(b)$ for some i and some $\lambda(b)$. b_i is on the common boundary of just $2m$ sets W_n, where m is the multiplicity of b_i. Then it is clear that there are just exactly $2m$ curves of $\tilde{G}[W]$, exactly one in each of these $2m$ sets, which have b_i as a limit point in one direction. For, if W_n has b_i on its boundary, then in the homeomorphism $\tilde{k}_n \colon W_n \to R_1$ the point b_i will map onto a point $(a, 0)$ and the inverse image of the line $x = a$ is the single curve of $\tilde{G}[W_n]$ which has b_i as a limit point. It follows at once that b_i is a branch of multipilicity $2m$ of $\tilde{G}$. Hence we have established that $\tilde{G}$ is a branched regular curve family with the same branch points as F. Again, just as above, it is clear that it is possible to find an arbitrarily small neighborhood U of each b_i which is homeomorphic to $|z| < 1$, and moreover, with a homeomorphism k carrying $F[\bar{U}]$ onto the level curves of $\Re(z^m)$ and $\tilde{G}[\bar{U}]$ onto the level curves of $\Im(z^m)$.

Finally, to complete the proof that $\tilde{G}$ is complementary to F, we note that by Corollary 2 to Theorem I 3. 5-3 we have at once that every curve of $\tilde{G}$ is a cross-section of F. This completes the proof of the following:

THEOREM 1. 2-1. *Every branched regular curve family F has at least one complementary family G $(= \tilde{G})$ as described above.*

1. 3. The fundamental theorem. Given any branched regular curve family F on π, we have shown the existence of a complementary family G; and also, we have shown [I] that each of these families is the level curve family of a continuous function (without relative extrema) $f(p)$ and $g(p)$ respectively. This enables us to define a single-valued mapping T_1 from the plane π to the complex w-plane as follows: $T_1(p) = u + iv$ where $u = f(p)$

and $v = g(p)$. $T_1(p)$ is clearly continuous, because f and g are continuous. Moreover, T_1 is locally a homeomorphism on R and is exactly m-to-1 in the neighborhood of an m-th-order branch point. To show this, it is sufficient to consider the special neighborhoods mentioned in the proof of the previous theorem, i. e., for every regular point we consider only a neighborhood U such that there is a homeomorphism of U onto the rectangle R_1 of the xy-plane such that $F[U]$ goes onto the lines $y = \text{constant}$ and $G[U]$ onto the lines $x = \text{constant}$. Then T_1 becomes a map of R_1 onto a rectangle in the uv-plane carrying the lines $y = \text{constant}$ onto $u = \text{constant}$ and $x = \text{constant}$ onto $v = \text{constant}$. It is clearly a homeomorphism since it is monotone on each line $x = \text{constant}$ and each line $y = \text{constant}$. This is exactly as in [III]. It is equally easy to show that in a neighborhood V of a branch point, where $F[V]$ and $G[V]$ map onto $\Re(z^m)$ and $\Im(z^m)$ respectively under a homeomorphism of V onto $|z| < 1$, T_1 carries V onto an open set and is at most m-to-1, where m is the multiplicity of the branch point (cf. [III]). Hence T_1 is not only interior but light (since for every point there is a neighborhood in which f and g take on the same value only a finite number of times in the neighborhood). It follows from Stoilow [V, Chapter V, part III, §5] and Whyburn [VII] that T_1 is topologically equivalent to an analytic function $W = \phi(z)$, i. e., there exists a homeomorphism $p = h(z)$ of the plane π onto either the domain $D_1 = \{z \mid |z| < 1\}$ or $D_\infty = \{z \mid |z| < \infty\}$ of the z-plane such that $\phi(z) = T_1[h(z)]$ is analytic. The family F' of level curves of the real part of $\phi(z)$ are just those curves mapping onto the lines $u = \text{constant}$ of the w-plane and hence are homeomorphic to F under h. It is thus proved that:

THEOREM 1. 3-1. *Given any branched regular curve family F there exists a function harmonic in either the finite plane or the unit circle whose level curves are homeomorphic to F.*

Since, if the function $u(x, y)$ is harmonic in a domain D, it is differentiable in D, its level curves will satisfy the differential equations $dx/dt = u_y$, $dy/dt = -u_x$, we have at once:

THEOREM 1. 3-2. *Given any branched regular curve family F, then there is a solution family of a system of differential equations to which it is homeomorphic.*

2. Decomposition of F into Half-Parallel Subfamilies.

2.1. Extended cross-sections.

THEOREM 2.1-1. *Let p be any regular point of π, C_p the curve of F through p, and let C be a curve containing a point q such that there is a cross-section pq. Then there will be a cross-section from p to an arbitrary point q' of T_C if and only if $q' \varepsilon C^*$, where C is directed so that $p \varepsilon \mathfrak{D}^*(C)$. Moreover, if $q' \varepsilon C^*$ and $U(qq')$ is any r-set (in $\mathfrak{D}^*(C)$) of qq', we may choose the cross-section qq' as follows: $qq' \equiv qrq'$ where qr lies on pq and rq' is in $U(qq')$.*

Proof. Suppose q' to lie on C^* and let $U(qq')$ be any r-set of qq'. Now moving along pq from p, the cross-section pq lies entirely inside $U(qq')$ from some point on, so we may choose some r on pq, with rq interior to U, letting prq now denote pq. We direct C_r so that $pr \subset \mathfrak{D}^*(C_r)$ and $rq \subset \mathfrak{D}^\#(C_r)$, which we can do by Theorem I 3.5-3 since prq is a cross-section. We replace rq by a cross-section rq' in U which is found as follows: U by definition is homeomorphic to the rectangle R_1 in the xy-plane, and we join in R_1 the image of r to that of q' by a straight line, whose inverse image we then take for rq'. Since the straight line is a cross-section of the image of F, i.e., the lines $y = k$, rq' will be also a cross-section, and will lie in the same domain $\mathfrak{D}^\#(C_r)$ as rq, since each cross the same curves in U. Hence, by Theorem I 3.5-3, we know that prq' is a cross-section.

It remains only to prove that if C' is any curve of T_C not on C^*, then there is no cross-section to C' from p. Now p lies in $\mathfrak{D}^*(C^*)$ and C' in $\mathfrak{D}^\#(C^*)$, hence any such cross-section, if it existed, would have to cross C^* and thus would have two points on T_C, contrary to the assumption that it is a cross-section.

THEOREM 2.1-2. *Let the trees of F be numbered as in* [I, *Section* 4], *i.e., in a standard numbering, using the concentric circles K_n of center p and radius n; further, let the cuts $\mathfrak{J}$ be removed from F, leaving $F^* = F[R^*]$. Then, outside every circle K_n lies at least one curve of F^* which can be reached from p by a cross-section lying in $R^* \cap \mathfrak{D}^*(C_p)$.*

Proof. Denote by $\{C\}$ the collection of all curves in $\mathfrak{D}^*(C_p)$ which can be reached by a cross-section from p lying in $R^* \cap \mathfrak{D}^*(C_p)$. We direct each curve of $\{C\}$ so that $\mathfrak{D}^*(C) \subset \mathfrak{D}^*(C_p)$. The existence of a cross-section from p to $q \varepsilon C$ makes this possible, i.e., direct C so that $\mathfrak{D}^\#(C) \supset pq$. $\{C\}$ will certainly not be empty since we assume p to be a regular point.

Now define on the curves of $\{C\}$ the positive real-valued function $d(C) = \underset{x \varepsilon C}{\text{g. l. b.}} \{\text{distance from } x \text{ to } p\}$. We have at once that C is outside K_n if and only if $d(C) > n$. Also it is clear that $\mathfrak{D}^*(C) \supset \mathfrak{D}^*(C')$ implies that $d(C) < d(C')$. To prove the theorem we must show that the numbers $d(C)$ are unbounded. We asume that this is not so; then there is a least upper bound d' of $d(C)$ for C in $\{C\}$. To show that this is impossible we shall choose $N > d'$ and consider intersections of curves of $\{C\}$ with K_N. Since we have assumed $d(C) \leqq d' < N$, every curve of $\{C\}$ will intersect K_N, although by Theorem I 4. 1-1 only a finite number of these curves lie completely inside K_N. All but a finite number of curves of $\{C\}$ in fact, not only have both endpoints outside K_N, but are themselves the only curve of T_C intersecting K_N. Hence, we may choose an infinite sequence of curves C_m of $\{C\}$ such that $d(C_m) \to d'$, $T_{C_m} \cap K_N \equiv C_m \cap K_N$, and $C_m \cap K_N$ contains neither endpoint of C_m. Having chosen such a sequence we find a subsequence q'_n of points $q'_n \varepsilon C_n$ which approach a regular point q as a limit and all lie on one side of the image of C_q in an r-neighborhood $U(q)$ (i. e., in the upper or lower half of R_0, the image of $U(q)$). This may be done as follows: first, by compactness of K_N we may find $q_{m_i} \varepsilon C_{m_i} \cap K_N$ (a subsequence of the m's) which converges to some point q'. Second, if q' is a regular point, we let $q = q'$ and choose a subsequence q'_n of the q_{m_i}'s, all of whose points lie in one side only of $U(q)$. Or, if q' is a branch point, let $V(q')$ be any admissible neighborhood of q'; then an infinite subsequence of the q_{m_i}'s will lie in one sector of V. If q is any regular point on either of the adjacent curves bounding this sector of V, there will be a corresponding sequence of points q'_n on the same curves C_n which contain the points q_n and such that $q'_n \to q$. The q'_n will lie on the same side of C_q in any r-neighborhood of q and is thus the desired sequence. Finally, we may choose a subsequence of q'_n which we will denote by r_n such that if qs is a cross-section from q to s in $U(q)$, where s lies on the same side of $U(q)$ as the q'_n, then the intersections $C_n \cap qs$ tend monotonely to q on qs (C_n denoting the curve on which r_n lies). Thus we have $d(C_n) \to d'$ monotonely since $\mathfrak{D}^*(C_n) \supset \mathfrak{D}^*(C_{n+1}) \supset C_q$ for all n. We direct C_q so that $\mathfrak{D}^*(C_n) \supset \mathfrak{D}^*(C_q)$.

Now choose in $\mathfrak{D}^\#(C_q)$ an r-set W of qq'' where q'' is any point of $C^\#_q$ which is in R^*. W is chosen so that its interior lies in R^*, which is possible by Theorem I 4. 1-4. Now for $n \geqq n_0$, r_n will lie in W, and since we have $\mathfrak{D}^*(C_{n_0}) \supset C^\#_q$ and $\mathfrak{D}^\#(C_{n_0}) \supset C_p$, we may extend the cross-section $pr_{n_0} \subset R^* \cap \mathfrak{D}^*(C_p)$ to a cross-section $pr_{n_0}q'' \subset R^* \cap \mathfrak{D}^*(C_p)$ by merely adding to it that cross-section $r_{n_0}q''$ in $W \cap \mathfrak{D}^*(C_{n_0})$ which is the inverse image in W of the straight line joining the images of r_{n_0} and q'' in R_1, the image of W. This

will be a cross-section by Theorem I 3. 5-3. Now since q'' is a regular point of a curve $C_{q''}$, if we take its direction such that $C^{\#}_{q''} \equiv C^{\#}_{q}$, we have $\mathfrak{D}^{\#}(C_{q''}) \supset C_p, C_n$; and $\mathfrak{D}^*(C_{q''}) \subset \mathfrak{D}^*(C_n)$ for all n, whence $d(C_{q''}) \geqq d'$. Now it is easy, however, by taking an r-neighborhood of q'' (which will lie in R^*) to extend $pr_{n_0}q''$ to a slightly larger cross-section $pr_{n_0}q''s$, and since $C_s \subset \mathfrak{D}^*(C_{q''})$, we have at once that $\mathfrak{D}^*(C_{q''}) \supset \mathfrak{D}^*(C_s)$, where C_s is directed as a curve of $\{C\}$, i. e. so that $\mathfrak{D}^*(C_s) \subset \mathfrak{D}^*(C_r)$. Hence $d(C_s) > d(C_{q''}) \geqq d'$. This is contrary to the assumption that d' is a bound of $d(C)$. Hence $d(C)$ is unbounded, which is what was to be proved.

By an *extended cross-section,* we shall mean any curve in $R = \pi - B$ which meets each curve of F at most once and tends to infinity in one or both directions. An extended cross-section is said to *tend properly to infinity* in R in a given direction on it, if it tends to infinity in that direction in such a way that the curves meeting it tend uniformly to infinity with their intersection points with the cross-section. We shall also speak of an *extended cross-section* in R^* which will be an extended cross-section as above, and lie entirely in $R^* = \pi - \tilde{\mathfrak{J}}$, i. e., it meets only curves of F^*.

THEOREM 2. 1-3. *If p is any regular point on a curve C of F^*, then there is an extended cross-section in R^* from p, which lies in $\mathfrak{D}^*(C_p)$ and tends properly to infinity.*

Proof. We consider a curve C in F^* and p any point on it. As before K_n will denote a circle with center at p and radius n; and for any point s we shall let $Q_n(s)$ denote a circle with center at s and radius so chosen that $Q_n(s)$ contains K_n. Now we choose a regular curve C_1 in $\mathfrak{D}^*(C_p) \cap R^*$ for which there is a cross-section pq_1 in $\mathfrak{D}^*(C_p) \cap R^*$ from p to q_1 on C_1. Direct C_1 so that $\mathfrak{D}^*(C_p) \supset \mathfrak{D}^*(C_1)$ and choose in $\mathfrak{D}^*(C_1) \cap R^*$ a curve C_2 outside of $Q_1(q_1)$ and such that a cross-section q_1q_2 in $\mathfrak{D}^*(C_1) \cap R^*$ exists with q_2 on C_2. Having chosen C_n and $q_n \varepsilon C_n$ in this manner, we choose for C_{n+1} any regular curve outside of $Q_n(q_n)$ for which there is a cross-section q_nq_{n+1} in $\mathfrak{D}^*(C_n) \cap R^*$ to q_{n+1} on C_{n+1}. We direct C_{n+1} so that $\mathfrak{D}^*(C_n) \supset \mathfrak{D}^*(C_{n+1})$. We continue this process indefinitely by Theorem 2. 1-2. Then the curves $pq_1, pq_1q_2, pq_1q_2q_3 \cdots$ will all be cross-sections by Theorem I 3. 4-5. They approach a curve Γ extending from p to infinity in $\mathfrak{D}^*(C_p) \cap R^*$ which is an extended cross-section extending from p to infinity in R^*. The curves intersecting Γ tend uniformly to infinity with any sequence of their points of intersection tending to infinity on Γ; since if r on Γ is beyond q_n, then C_r lies outside K_n. Thus Γ is an extended cross-section tending properly to infinity in R^*.

2. 2. Half-parallel subfamilies of F. We mean by a *half-parallel subfamily of F* the collection of all curves of F which intersect an extended cross-section Γ tending from a point p on a curve C_p properly to infinity. And we shall mean by a *complete* half-parallel subfamily of F the curve $C_p{}^*$ together with all curves of F crossing Γ (C_p being so directed that $\mathfrak{D}^*(C_p) \supset \Gamma$). The first of these sets is homeomorphic to the lines $y = k$, $k \geqq 0$ of the half-plane by Theorem 1. 1-1 and the same reasoning as used in the proof of that theorem will establish this homeomorphism for the second case, the complete half-parallel subfamily also. The first will be denoted by S and the second by S^*. Clearly $S^* \supset S$ and when C_p is a regular curve they are identical. C_p is called the *initial curve* of S, $C_p{}^*$ the initial curve of S^*.

If $\Gamma(q)$ is any half-open cross-section of F tending from a regular point q properly to infinity, then the boundary of $S(\Gamma)$, $S(\Gamma)$ being the collection of curves intersecting Γ, is best described in terms of maximal chains C^*, $C^\#$ and the sets $\delta(C+)$, $\delta(C-)$ defined in [I, Section 3]. We shall refer to these latter two sets as *mixed maximal chains,* since they consist of two subchains of maximal chains, one clockwise adjacent, the other counterclockwise adjacent, e. g., $\delta(C+) = \delta^*(C+) \cup \delta^\#(C+)$ (which may be empty). $\delta(C)$ will denote $\delta(C+) \cup \delta(C-)$; it is empty if and only if C is a regular curve.

THEOREM 2. 2-1. *The boundary of $S(\Gamma)$ is a collection of maximal chains C^*, $C^\#$ and mixed maximal chains $\delta(C)$, where $\delta(C)$ is on the boundary if and only if C is in $S(\Gamma)$. From each set T_C of F there is either* (1) *no point,* (2) *exactly one maximal chain, or* (3) *a set $\delta(C)$ of T_C on the boundary of $S(\Gamma)$.* (1), (2) *and* (3) *are mutually exclusive.*

Proof. Suppose $C \,\varepsilon\, S(\Gamma)$ is a singular curve, then $\delta(C)$ is in the boundary of $S(\Gamma)$, for if we consider any point q on $\delta(C)$ there exists an r-set $U(pq)$ containing q and $p = C \cap \Gamma$ (since C lies on an adjacent chain with C_q); choosing a sequence of points $p_n \to p$, $p_n \,\varepsilon\, U \cap \Gamma$, we can find by Theorem I 3. 5-2 a sequence $q_n \,\varepsilon\, U$ such that $q_n \,\varepsilon\, C_{p_n}$ for all n and $q_n \to q$. Whence q is a limit point of points of $S(\Gamma)$. But, if q is in $\delta(C)$, it is on a curve of T_C other than C; and C_q therefore cannot intersect Γ. Hence C_q is not in $S(\Gamma)$, and thus q is on the boundary of $S(\Gamma)$. Moreover, no other curves of T_C can in this case be on the boundary of $S(\Gamma)$, for $S(\Gamma)$ is clearly contained in $\mathfrak{D}^*(C) \cup C \cup \mathfrak{D}^\#(C)$, a complementary domain of $\delta(C)$, whereas every other curve of T_C lies in one or two other complementary domains of $\delta(C)$. (Note: $\delta(C)$ divides π into at most three Jordan domains.

On the other hand, suppose that C is a curve of F on the boundary

of $S(\Gamma)$. (Note: from what follows it is clear that the boundary is indeed a union of curves of F.) Then, directing C so that $\mathfrak{D}^*(C)$ contains the initial point of Γ, we note that if p is a point on C, limit point of a sequence p_n of $S(\Gamma)$, then there is an r-set $U(pq)$ of any arc pq on C^* and a sequence $q_n \to q$ with $q_n \varepsilon C_{p_n}$, $C_{p_n} \subset S(\Gamma)$, from which we conclude that q is either in $S(\Gamma)$ or on its boundary. If C^* does not cross Γ, then q will be on the boundary and C^* is a boundary curve of $S(\Gamma)$. When this is the case, C^* divides π into two domains $\mathfrak{D}^*(C^*) \supset S(\Gamma)$ and $\mathfrak{D}^\#(C^*) \supset T_C - C^*$, whence no other points of T_C other than those of C^* are on the boundary of $S(\Gamma)$. But, if C^* crosses Γ at a point p on a curve C', then we are back in the previous case and $\delta(C') = [C^* \cup C^\#] - C'$ is the boundary in T_C of $S(\Gamma)$.

THEOREM 2.2-2. *Let q be a point on a curve C_q of $F^* = F[R^*]$ and let $\Gamma(q)$ be a cross-section tending properly to infinity in R^* in each direction. Further, let h be any homeomorphism of R^* onto the xy-plane, then $h[\Gamma(q)]$ is a cross-section of the family $h[F^*]$ (filling the xy-plane) which tends properly to infinity in both directions on the xy-plane.*

Proof. On the xy-plane we let K_n denote a circle of radius n, center $h(q)$. We must show that for every n there are points q'_n, r'_n on $\Gamma' = h[\Gamma(q)]$ such that every curve of $h(F^*)$ intersecting Γ' at points outside the arc $q'_n r'_n$ will lie outside K_n. If this is not the case, as we shall assume, we will be able to find a sequence of points t'_n receding to infinity on Γ' such that each $C_{t'_n}$ intersects a fixed circle K_N of the circles K_n. Now the inverse image K of K_N is a simple closed curve in R^* containing q in its interior. We will denote by C_n the inverse image of $C_{t'_n}$ and by t_n the inverse image of t'_n. Every C_n must then intersect K and hence intersect some circle with center at q which contains K. But this contradicts the assumption that $\Gamma(q)$ tended properly to infinity in R^*, since we have a sequence t_n approaching infinity on $\Gamma(q)$, but the curves C_{t_n} do not approach infinity. Hence the theorem must be true.

W. Kaplan introduced the notion of admissible collections of finite sequences in order to number the half-parallel subsets of a regular curve family filling an open simply connected domain. The concept is so similar to that already considered in the numbering of curves of a tree that we shall be able to use the same notation as in that section. As in Kaplan [II], we shall call a collection A of finite sequences *admissible* if

(1) A contains the one-element sequence 1 and no other one-element sequences, and

(2) $\alpha, k \,\varepsilon\, A$ implies $\alpha, k-1 \,\varepsilon\, A$ if $k > 1$ and implies $\alpha \,\varepsilon\, A$ if $k = 1$.

Now, if we have a regular curve family F' filling the xy-plane, and if we have assigned to each point (x,y) an extended cross-section $\Gamma(x,y)$ tending properly to infinity in both directions, then for any fixed curve C_1 it was shown in [II] that we can decompose $F'[C_1 \cup \mathfrak{D}^*(C_1)]$ into a collection of non-overlapping, half-parallel subfamilies $S(\alpha)$ which will be numbered by the finite sequences $\{\alpha\}$ of an allowable collection A. Each half-parallel family $S(\alpha)$ will be the set of all curves intersecting a cross-section $\Gamma(\alpha)$ tending from an initial curve C_α to infinity and lying on some $\Gamma(x,y)$ as chosen above; C_α will be the only curve of $S(\alpha)$ mapped onto the x-axis in the homeomorphism of $S(\alpha)$ onto the lines $y = k \geqq 0$ and the complete boundary of $S(\alpha)$ will be, in addition to C_α, just exactly the curves $C_{\alpha,k}$. Note that when we write C_α we mean to indicate that C_α is an initial curve of some $S(\alpha)$ in the decomposition of F', whereas $C(\alpha)$ will, as in [I], indicate that C is the curve of a numbered tree which has been assigned the signed sequence α in the numbering of the tree.

As a corollary to the preceding Theorem 2.2-2 plus the proof of the facts mentioned in the preceding paragraph from [II], we can immediately state the following theorem:

THEOREM 2.2-3. *Given the family $F^* = F[R^*]$ and an arbitrary regular curve C_1 of F^*, we can decompose $F^*[C_1 \cup \mathfrak{D}^*(C_1)]$ (which is the same as $F[C_1 \cup \mathfrak{D}^*(C_1) \cap R^*]$) into a collection of non-overlapping half-parallel subsets $S(\alpha)$, each $S(\alpha)$ being all curves intersecting a cross-section $\Gamma(\alpha)$ tending from a curve C_α in F^* properly to infinity in R^*.*

In order to study the relation between an arbitrary tree T of F and a given decomposition of F^* into sets $S(\alpha)$ ($\alpha \varepsilon A$, as described above), it is convenient to adopt some new notation. $A(T)$ will denote the subset of A containing all sequences α such that $S(\alpha) \cap T \neq 0$; and $A_n(T)$ the subset of all sequences of $A(T)$ of order n. We denote by $N(T)$ the smallest integer n such that $A_n(T)$ is not empty. It is clear that $\Gamma(\alpha)$ can have at most one point on T, and hence $S(\alpha) \cap T$ is a curve of F^* or is empty. If $\Gamma(\alpha) \cap T$ is the initial point of $\Gamma(\alpha)$ we say that $\Gamma(\alpha)$, or $S(\alpha)$, *begins at* T; in this case $C_\alpha = S(\alpha) \cap T$. When $\Gamma(\alpha) \cap T$ is a point of $\Gamma(\alpha)$ other than the initial point, then $\Gamma(\alpha)$, or $S(\alpha)$ is said to *straddle* T. In the former case $S(\alpha)$ lies in one complementary domain of T, in the latter in two. Using these notations, we may state the following properties:

(1) If α, β are *distinct* elements of A with $\beta \varepsilon A(T)$, and α either an element of $A(T)$ or such that points of T lie on the boundary of $S(\alpha)$; then $S(\alpha)$, $S(\beta)$ cannot each intersect the same complementary domain of T.

(2a) If $A_N(T)$, $N = N(T)$, has one element α, then either $S(\alpha)$ straddles T, or if $S(\alpha)$ begins at T, then $C_\alpha^* \cap R^* = S(\alpha) \cap T$, i. e., C_α^* has just *one* curve in R^*.

(2b) If $A_N(T)$ has more than one element, then every element of $A_N(T)$ is of the form β, k for fixed β of order $N - 1$ and $C_{\beta,k}$ for $\beta, k \varepsilon A_N(T)$ are just those curves of a maximal chain C^* which are in R^*.

(3) Let γ be an element of $A_{N+k}(T)$, then every lower segment of γ of order $\geqq N(T)$ is in $A(T)$, i. e., for $0 \leqq j \leqq k$ we have $\gamma_{N+j} \varepsilon A_{N+j}(T)$, where, as previously, γ_{N+j} is the sequence consisting of the first $N + j$ elements of the sequence γ.

(4) A necessary condition that $S(\alpha)$ straddle T is that $\alpha \varepsilon A_N(T)$ and is the only element of $A_N(T)$.

First we prove (1). Let $\mathfrak{D}^*(C)$ be a complementary domain of T, bounded by C^* on T. Suppose that $S(\alpha)$ and $S(\beta)$ both have points in $\mathfrak{D}^*(C)$. Then there is a point p_1 on $\Gamma(\alpha)$, p_2 on $\Gamma(\beta)$, each in $\mathfrak{D}^*(C)$. Now since $\beta \varepsilon A(T)$, $\Gamma(\beta)$ has a point q_2 on C^* and p_2q_2, an arc on $\Gamma(\beta)$, lies in $\mathfrak{D}^*(C) \cup C^*$. In either of the possibilities for α mentioned above, there would be a point q_1 on C^* which was a limit point of points q'_n in $S(\alpha)$. If $\alpha \varepsilon A(T)$ then q'_1 and q'_n may be taken on $\Gamma(\alpha)$, otherwise q'_n will consist of points in $\mathfrak{D}^*(C)$ which are in $S(\alpha)$ but not on $\Gamma(\alpha)$. It follows by arguments used many times above, i. e. considering an r-set of q_1q_2, etc., that there is a cross-section from q_1 on C^* into $\mathfrak{D}^*(C)$, which may be shown to cross a curve also crossed by p_2q_2. This curve would have to be in both $S(\alpha)$ and $S(\beta)$ which is impossible since α, β were assumed distinct. The following lemma will be useful in proving properties 2-4.

LEMMA. *If $\alpha \varepsilon A(T)$ and $\alpha, k \not\varepsilon A(T)$, then no sequence γ of $A(T)$ can have α, k as a lower segment.*

Proof. $C_{\alpha,k}$ lies on the boundary of $S(\alpha)$ but is not in T, nor is any curve of $S(\alpha, k)$ in T by hypothesis. Assuming $C_{\alpha,k}$ directed so that $S(\alpha, k) \subset \mathfrak{D}^*(C_{\alpha,k})$, we have by Theorem 2. 2-1 two possibilities: (a) the entire curve $C^{\#}_{\alpha,k}$ is on the boundary of $S(\alpha)$ and is all of this boundary in the tree $T' = T_{C_{\alpha,k}}$, or (b) T' intersects $S(\alpha)$ on a curve C of T', and then $C_{\alpha,k} \subset \delta(C)$ where $\delta(C)$ is on the boundary of $S(\alpha)$ and is all of this boundary in T'. In case (a) every curve of $C^{\#}_{\alpha,k} \cap R^*$, being on the boundary of $S(\alpha)$, is a curve $C_{\alpha,k'}$ for some k'. We have $\mathfrak{D}^{\#}(C^{\#}_{\alpha,k}) \supset T$, since it contains $S(\alpha)$ which intersects T. In case (b) $\delta(C) = \delta(C+) \cup \delta(C-)$ divides π into three domains (or two if one of the sets $\delta(C \pm)$ is empty);

one of these which we denote D_1 contains C and hence $S(\alpha)$ and T. The others, D_2 and D_3, contain all other curves of T'. $\delta(C)$ is the complete boundary in T' of $S(\alpha)$, hence every curve of $\delta(C) \cap R^*$ is a curve $C_{a,k'}$ for some k'.

The remainder of the proof depends on the fact that $S(\beta) \cup S(\beta, k)$ is always a connected set. If there exists any sequence $\gamma = \alpha, k, n_1, \cdots, n_r$ such that γ is in $A(T)$ then, $S(\gamma)$ must clearly have points in $\mathfrak{D}^{\#}(C^{\#}_{a,k})$ above in case (a) or in D_1 in case (b), these being the domains of T' in which T lies, and hence, so also has the set $\Sigma = \bigcup_{j=0}^{r} S(\alpha, k, n_1, n_2, \cdots, n_j)$. Moreover, the set Σ is connected, and $S(\alpha, k)$ which is in this set lies in $\mathfrak{D}^*(C_{a,k})$ in case (a), and in D_2 or D_3 in case (b). Thus Σ has points on either $C^{\#}_{a,k}$ or $\delta(C)$, the boundary curve, i. e., for a $j \neq 0$ there is a curve of $C^{\#}_{a,k}$ or $\delta(C)$ as the case may be in $S(\alpha, k, n_1, \cdots, n_j)$. But each such curve as already pointed out is a curve $C_{a,k'}$, which is a contradiction.

The lemma implies in particular, that if α and $\alpha, n_1, \cdots, n_r \varepsilon A(T)$ then $\alpha, n_1, \cdots, n_j \varepsilon A(T)$, $j \leqq r$. Hence (3) will follow if we prove that every sequence of $A(T)$ contains a lower segment in $A_N(T)$.

Now we turn to an examination of the possibilities for $A_N(T)$ and completion of the proof of (3). Suppose that α is an element of $A_N(T)$. Then either $S(\alpha)$: (i) straddles T, or (ii) begins at T. In the former case let $C = S(\alpha) \cap T$, then $\delta(C)$ is the complete boundary of $S(\alpha)$ in T, and we know that every curve in R^* of $\delta(C)$ is in the collection $\{C_{a,k}\}$. Moreover, $\delta(C)$ divides π into three (or two) domains $D_1, D_2, (D_3)$ of which the first contains C_1, and of T, *only* the curve C. Now let γ be any sequence of $A(T)$. $S(\gamma)$ must, by (1), lie in D_2 or D_3. But $\bigcup_{i=1}^{n} S(\gamma_i)$ is a connected set containing C_1 (i. e., C_a, $\alpha = 1$), hence points of D_1 and also points of D_2 or D_3. It must then contain a curve $C_{a,k}$ of $\delta(C)$, and therefore $S(\alpha)$, i. e., α is a lower segment of γ. Since this is only possible if γ is of order $> N$ we conclude α is the only element of $A_N(T)$.

In the case (ii) where $S(\alpha)$ begins at T, we have C_a on the boundary of $S(\beta)$, where β is of order $N - 1$ and $\alpha = \beta, k$. In fact, $C^{\#}_a$ is the complete boundary on T of $S(\beta)$ (the curves of $C^{\#}_a \cap R^*$ are all in the set $\{C_{\beta,k'}\}$ for some k' and therefore are in $A_N(T)$); and we have $\mathfrak{D}^{\#}(C^{\#}_a) \supset S(\beta)$, $\mathfrak{D}^*(C^*_a) \supset S(\alpha)$. Now let us suppose that $\gamma \varepsilon A(T)$, then $S(\gamma)$ by (1) cannot lie in $\mathfrak{D}^{\#}(C^{\#}_a)$, hence must lie in $\mathfrak{D}^*(C^{\#}_a)$. But $\bigcup_{i=1}^{n} S(\gamma_i)$ is connected and has a point in $\mathfrak{D}^{\#}(C^{\#}_a)$, namely, any point of C_1. Thus this set has a

point on $C^{\#}_{\alpha} \cap R^*$ and hence contains a curve $C_{\alpha,k'}$. It follows that every sequence of $A(T)$ has a lower segment in $A_N(T)$. This proves (1) and completes the proof of (3).

To prove (4) we need show only that if $S(\alpha)$ straddles T, then no lower segment of α is in $A(T)$. If $\alpha = \beta, k$, so that β is the lower segment α_{n-1}, then if any lower segment of α is in $A(T)$, β is also by our lemma. Then C_α, being on the boundary of $S(\beta)$, we necessarily have $S(\beta)$, $S(\alpha)$ in different domains of T_{C_α}. This is impossible unless $T = T_{C_\alpha}$, for we would otherwise have points of T in two different domains of T_{C_α}. But then $S(\alpha)$ does not straddle T, for on the contrary $C_\alpha \subset T$.

As above we consider the branched regular curve family F with a regular curve C_1 of F and a decomposition of the corresponding $F^*[C_1 \cup \mathfrak{D}^*(C_1)]$ into sets $S(\alpha)$ with initial curves $C(\alpha)$. Then using properties 1-4, we have the following:

THEOREM 2. 2-4. *The complete half-parallel subfamilies $S^*(\alpha) = S(\alpha) \cup C^*_\alpha$ decompose $F[\mathfrak{D}^*(C_1) \cup C_1]$ into a family of half-parallel subsets which intersect only on a curve of their initial curves, i. e., $S^*(\alpha) \cap S^*(\beta) = 0$ or C where $C^* = C^*_\alpha$ and $C^{\#} = C^{\#}_\beta$.* (See Figure 4.)

Proof. First to prove that every curve of $F[C_1 \cup \mathfrak{D}^*(C_1)]$ is included in this decomposition we note that every curve of $F^*[C_1 \cup \mathfrak{D}^*(C_1)]$ is automatically included, being already in a set $S(\alpha)$ of the decomposition of that part of the simply connected region R^* included in $\mathfrak{D}^*(C_1)$. Thus we have only to consider curves of $\tilde{\mathfrak{J}}$; let C be a curve of $F[C_1 \cup \mathfrak{D}^*(C_1)]$ which is not in R^* and let T denote the tree which contains it. Then no cross-section $\Gamma(\alpha)$ has a point on C. C will be on the boundary of two distinct sets $S(\alpha)$ and $S(\beta)$ in $\mathfrak{D}^*(C)$ and $\mathfrak{D}^{\#}(C)$ respectively. They cannot coincide since if they did, then it would mean that $S(\alpha) = S(\beta)$ would straddle T, for otherwise the set $S(\alpha)$ lies in a single domain of T. But then, in this case, since the cross-section $\Gamma(\alpha)$ would have points in two domains both having C and only C as common boundary, it would have to contain a point of C, which is clearly impossible if C is not in R^*.

Now, if either α or β, say α, is of order $> N(T)$ then, since C, a curve of T is on the boundary of $S(\alpha)$, α, k for some k is in $A(T)$. Hence by (3), α must also be in $A(T)$. Then by (4), C_α must lie on T, whence we have at once that $C^*_\alpha = C^* \supset C$ and hence C is in $S^*(\alpha)$. Thus it remains to show that either α or β must be of order $> N$. Clearly each is of order $\geqq N - 1$ since, for example, curves of T are on the boundary of $S(\alpha)$ hence either $\alpha \varepsilon A(T)$ or $\alpha, k \varepsilon A(T)$ for some k. Assume α is of order $N - 1$, then $\Gamma(\alpha)$ does not have a point on T and by Theorem 2. 2-1 all of C^* is on

the boundary of $S(\alpha)$, and every curve of $C^* \cap R^*$ is in the set $\{C_{\alpha,k}\}$, and for each of these $\alpha, k \, \varepsilon \, A_N(T)$. Now, since β, k' for some k' is in $A(T)$, β is of order $\geqq N-1$. If now β were of order $N-1$, then there must be a $\beta, k' \, \varepsilon \, A_N(T)$ and β, k' would for some k equal α, k by property (2b), whence $\beta = \alpha$ by (1); or if β were of order N, then it is of the form α, k for some k. This latter would mean that the common boundary of the

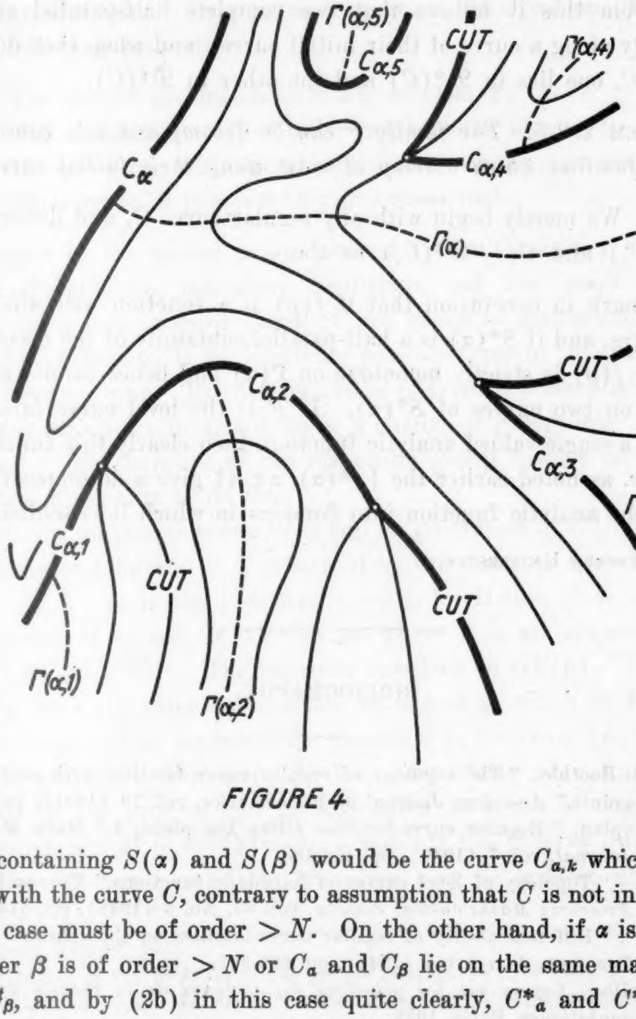


FIGURE 4

domains containing $S(\alpha)$ and $S(\beta)$ would be the curve $C_{\alpha,k}$ which must then coincide with the curve C, contrary to assumption that C is not in R^*. Hence β in this case must be of order $> N$. On the other hand, if α is of order N, then either β is of order $> N$ or C_α and C_β lie on the same maximal curve $C^\#_\alpha = C^\#_\beta$, and by (2b) in this case quite clearly, C^*_α and C^*_β could not have a boundary curve C in common. Hence either α or β is of order $> N$ and we have already shown that in this case C is in the initial curve C^*_α or C^*_β of either $S^*(\alpha)$ or $S^*(\beta)$ respectively.

Next it must be shown that if C_α is the initial curve of a set $S(\alpha)$, then

for any $S(\beta)$ which intersects C^*_α, the intersection must be C_β. Let C be the curve of intersection, i. e., $C = C^*_\alpha \cap S(\beta)$. Thus, $\alpha, \beta \in A(T)$ where T is the tree containing C_α. Now $S(\alpha)$ and $S(\beta)$ cannot have points in the same complementary domain of T by (1), which means in particular that $S(\beta)$ cannot straddle T, i. e., $\Gamma(\beta) \cap C =$ initial point of $\Gamma(\beta)$, since one complementary domain of C is $\mathfrak{D}^*(C^*_\alpha)$, C lying as it does on C^*_α, and $S(\alpha)$ lies in this complementary domain. Hence $C_\beta = C$ which was to be proved. From this it follows that two complete half-parallel subfamilies intersect only along a curve of their initial curves, and when they do intersect on a curve C, one lies in $\mathfrak{D}^*(C)$ and the other in $\mathfrak{D}^\#(C)$.

THEOREM 2. 2-5. *The family F can be decomposed into complete half-parallel subfamilies which overlap at most along their initial curves.*

Proof. We merely begin with any regular curve C_1 and decompose both $C_1 \cup \mathfrak{D}^*(C_1)$ and $C_1 \cup \mathfrak{D}^\#(C_1)$ as above.

We remark in conclusion that if $f(p)$ is a function with the family F as level curves, and if $S^*(\alpha)$ is a half-parallel subfamily of the decomposition, then clearly $f(p)$ is strictly monotone on $\Gamma(\alpha)$ and hence cannot assume the same value on two curves of $S^*(\alpha)$. If F is the level curve family of the real part of a single-valued analytic function, then clearly this function is 1-1 in $S(\alpha)$, i. e. as noted earlier the $\{S^*(\alpha), \alpha \in A\}$ give a decomposition of the domain of the analytic function into domains in which it is schlicht.

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY.

[I] W. M. Boothby, "The topology of regular curve families with multiple saddle points," *American Journal of Mathematics*, vol. 73 (1951), pp. 405-438.

[II] W. Kaplan, "Regular curve-families filling the plane, I," *Duke Mathematical Journal*, vol. 7 (1940), pp. 154-185.

[III] ———, "Topology of level curves of harmonic functions," *Transactions of the American Mathematical Society*, vol. 63, No. 3 (1948), pp. 514-522.

[IV] ———, "Differentiability of regular curve-families on the sphere," *Lectures in Topology*, Ann Arbor (1941), pp. 299-301.

[V] S. Stoilow, *Leçons sur les principes topologiques de la théorie des fonctions analytiques*, Paris, 1938.

[VI] H. Whitney, "Regular families of curves," *Annals of Mathematics* (2), vol. 34 (1933), pp. 244-270.

[VII] G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications, vol. 28, New York (1942).

FINITE METABELIAN GROUPS AND THE LINES OF A PROJECTIVE FOUR-SPACE.*

By H. R. Brahana.

1. Introduction. This paper presents a classification of the groups with the following properties:

(1) The groups are metabelian, i. e. of class 2;

(2) The elements are all of order p except the identity;

(3) In each group the central and coommutator subgroups coincide;

(4) Each group is generated by five elements.[1]

It was shown in the second paper just cited that there is a unique group of order p^{15} satisfying the above conditions, and that every other group satisfying them is a quotient group of this group of maximum order. Our method of classification is to determine possible types of invariant subgroup of the maximum group. The determination of types is accomplished by means of properties which are invariant under a change of generators. A geometric representation of subgroups of the maximum group enables us to make use of geometric invariants to do this.

Denote the maximum group by G and its generators by U_1, U_2, U_3, U_4, U_5. The commutator subgroup C is generated by the elements $c_{ij} = U_i^{-1}U_j^{-1}U_iU_j$, $i, j = 1, \cdots, 5$. c_{ii} is the identity; $c_{ji} = c_{ij}^{-1}$. Hence, C is of order p^{10}. Every element of G can be written cx where c is an element of C and $x = U_1^{x_1}U_2^{x_2}U_3^{x_3}U_4^{x_4}U_5^{x_5}$. The numbers x_i belong to $\mathrm{GF}(p)$. Two elements cx and $c'y$ have the same commutator as x and y, which is $\Pi c_{ij}^{p_{ij}}$, $i < j$, $p_{ij} = x_iy_j - x_jy_i$. If we represent the element x by the point $(x_1, x_2, x_3, x_4, x_5)$ in a finite projective four-space X, the commutator of x and y may be represented by the line xy and its expression in terms of commutators of pairs of generators of G will have for exponents the Plücker line-coordinates of xy. The Plücker coordinates of a line may be ordered and each line

* Received May 12, 1950.

[1] The relation of the problem of the groups having the first three properties to the general problem of groups of order p^m was discussed at considerable length in the author's paper "Metabelian groups and trilinear forms," *Duke Mathematical Journal*, vol. 1 (1935), pp. 185-197. Groups generated by four elements were classified in "Finite metabelian groups and Plücker line-coordinates," *American Journal of Mathematics*, vol. 62 (1940), pp. 365-379.

of the four-space may be represented by a point of a finite projective space S of nine dimensions. Every line in X determines a point in S but not every point in S corresponds to a line in X. Every point in S, however, determines a cyclic subgroup of C, necessarily of order p.

A set of defining relations of G contains the definitions of the c_{ij}'s given above and some other relations which were not given explicitly, viz.: $c' \cdot cx = cx \cdot c'$, which says G is metabelian; $cx \cdot c'y = c'y \cdot cx$ for all y only if $x_i = 0$, which says G is not the direct product of an abelian group and a metabelian group generated by fewer than five elements; and $\Pi c_{ij}^{\alpha_{ij}}$ is the identity only if $\alpha_{ij} = 0$ for all pairs i, j, which says that the c_{ij}'s are independent and G is the maximum group. Hence any group satisfying the conditions of the present problem will have defining relations the same as these except that the last will be replaced by one or more relations stating that one or more elements of the commutator subgroup is the identity. By a well-known theorem due to Dyck [2] this says that every group of the type we are considering is a quotient group of G with respect to some subgroup of C.

Every subgroup of C corresponds to some linear space in S, and every linear space in S determines a subgroup of C. Our problem then involves the determination of relations of all possible linear subspaces of S to the subset V of points of S which represent lines of X. A simple isomorphism of G with itself determines a collineation of the space X, a transformation of the space S into itself, and a transformation of V into itself. A simple isomorphism of G with itself which transforms U_i into $c^{(i)}U_i$, $i = 1, \cdots, 5$, determines the identity as the collineation in X and hence as the transformation in S; it determines the identity isomorphism also of the quotient group G/C. Isomorphisms of G with itself which are significant as far as S is concerned are those which determine non-identity isomorphisms of G/C with itself, and there is a one-to-one correspondence between these and the collineations of X. Two subgroups of C which are conjugate under the group of isomorphisms of G determine simply isomorphic quotient groups of G; their corresponding linear spaces in X are the same or are conjugate under a non-identity collineation of X. Hence we are interested in the conjugacy of linear spaces in S under collineations in X. These collineations may always be thought of as changes of generators of G/C.

In the following pages the types of point, line, plane, and three-space in S are given. For each type a canonical form is given. The canonical

[2] "Gruppentheoretische Studien," *Mathematische Annalen*, vol. 20 (1880), p. 14.

forms are all given for $p = 7$. One purpose of the canonical form is to establish the existence of the type; another purpose is to exhibit the relation to V of the corresponding subspace of S. These relations always come finally to the question of the reducibility or irreducibility of polynomials in one variable with coefficients in GF(p). The canonical form sometimes depends on the selection of a polynomial belonging to GF(7) having certain properties; the existence of a polynomial having the required properties nowhere depends on p being 7. For $p = 3$ or 5 some of the types may not exist or may coincide with others that are different from them for $p \geq 7$; the argument is carried out in geometric terms and in a few places might run into difficulties if a conic has only four points or if a pencil has only four or six members.

The grassmannian of lines in a four-space has been studied before, generally for the field of complex numbers. A paper [3] by J. A. Todd develops a wealth of geometry and gives many references to the literature. The interpretation of these results in a geometry based on GF(p) is in general obvious. In the third section of his paper Todd deals with linear spaces in S, with emphasis on their intersections with V. He determines points, lines, and planes; he gives some results about five-spaces which can be turned into theorems about three-spaces by means of the existing duality. The completeness of the list of planes which follows, and which was obtained independently, gives a verification of the completeness of Todd's list, which he does not state explicitly.

The completeness of the list of three-spaces which follows is not certain. Every S_3 in S determines a quintic polynomial in a single variable x. If the quintic is reducible, a coordinate system in X can be selected so that S_3 has one of the forms 1 to 53. These 53 types are distinguished from one another and from S_3's which give irreducible quintics by the presence in them of special points, lines, and planes. If the quintic is irreducible none of these special points, lines, or planes is present. Every irreducible quintic determines an S_3. The irreducible quintics belong to sets of conjugates under rational transformations of the variable x. An S_3 given by an irreducible quintic $f(x)$ may, by a change of coordinates in X, be made to give any quintic conjugate to $f(x)$, or it may be made to give a quintic which is not conjugate to $f(x)$; it is not certain that $f(x)$ can be changed to an arbitrary irreducible quintic $g(x)$. The type 54 can be characterized, and in many ways, by geometric properties which determine the coordinate system. An S_3

[3] "The locus representing the lines of four-dimensional space and its application to linear complexes in four dimensions," *Proceedings of the London Mathematical Society*, series 2, vol. 30 (1921), pp. 513-550.

which gives an irreducible quintic has determined in it by its relation to V a set of p^2+1 rational cubic curves no two of which intersect; the points of the cubics therefore fill S_8. The form 54 is determined by the presence in the set of two cubics properly related to each other. The pair is determined by either of the cubics. The relations that two cubics can have to each other fall into at most 21 sets, irrespective of p and irrespective of the irreducible quintic. It has been found that not every cubic can be a member of a "canonical" pair; it seems likely that every such set of cubics contains at least one canonical pair.

2. Some relations between elements in X and elements in S. In order to avoid double subscripts we adopt a notation as follows: let x and y be two points of the four-space X; then the line xy determines in S the point

$$(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}) = (p_{12}, p_{13}, p_{14}, p_{15}, p_{23}, p_{24}, p_{25}, p_{34}, p_{35}, p_{45}).$$

The manifold V is the set of points in S whose coordinates satisfy certain obvious relations among the two-rowed determinants p_{ij}. In terms of the a's they are $B_1=0, \cdots, B_5=0$, if $B_1, \cdots, B_5$ denote

$$a_1a_8 - a_2a_6 + a_3a_5, \quad a_1a_9 - a_2a_7 + a_4a_5, \quad a_1a_{10} - a_3a_7 + a_4a_6,$$
$$a_2a_{10} - a_3a_9 + a_4a_8, \quad a_5a_{10} - a_6a_9 + a_7a_8,$$

respectively. Three of the five relations are independent, but all five are linearly independent.

Two lines of X determine two points of V. If the two lines of X intersect, it is well known [4] that the lines of the pencil defined by them are represented by the points of a line on V, and the lines in the plane determined by them are represented by the points of a plane on V.

There are obviously two kinds of points in S: (1) points on V; and (2) points not on V. Points on V are all alike since each represents a line X and there exists a collineation of X transforming any line into any other.

We proceed to prove that points of S not on V are also all of one type. First we prove the theorem:

Every point of S not on V is on a line joining two points of V.

Let $P=(a_1, \cdots, a_{10})$ be not on V. We may suppose that $a_{10} \neq 0$. Let us denote the vertices of the frame of reference in X by $A_1=(1,0,0,0,0)$, $A_2=(0,1,0,0,0), \cdots, A_5=(0,0,0,0,1)$. If we change the coordinate

---

[4] The two lines in X are in a three-space in X and we apply Theorem 30, Veblen and Young, *Projective Geometry*, vol. 1, p. 329.

system in X to one that has the same vertices except that $A'_4 = (a_4, a_7, a_9, a_{10}, 0)$ and make the corresponding changes in coordinates in S, then P will become $(a'_1, a'_2, a'_3, 0, a'_5, a'_6, 0, a'_8, 0, 1)$. This is a point of the line joining $(a'_1, a'_2, 0, 0, a'_5, 0, 0, 0, 0, 0)$ and $(0, 0, a'_3, 0, 0, a'_6\ 0, a'_8, 0, 1)$, and both of these points are on V. These points represent lines in X which are skew, since P is not on V. We may select a new frame of reference in X with two vertices on each of these two lines. If we take A_1 and A_2 on one and A_3 and A_4 on the other, the point P becomes $(1, 0, 0, 0, 0, 0, 0, 1, 0, 0)$.[5]

We have shown the existence of a line on two points Q and Q' of V through any point P not on V. Q and Q' are images in S of two lines q and q' in X. The two skew lines q and q' determine a three-space in X. A second line on P which intersects V in two points determines a three-space in X. We proceed to prove that these two three-spaces coincide. We suppose P to have the form just obtained; we take a point $Q = (a_1, a_2, \cdots, a_{10})$ on V, and find the condition on the a's in order that the line PQ meet V in two distinct or coincident points. Expressing the coordinates of a point on PQ in parametric form in terms of the coordinates of P and Q and substituting in the equations of V we obtain the conditions for two solutions to be that $a_4 = a_7 = a_9 = a_{10} = 0$. This says that Q must be a point on V which corresponds to a line in the three-space $x_5 = 0$ of X. Hence,

Every point of S not on V determines a unique three-space R in X. The lines of this three-space are the only lines in X such that their corresponding points on V can be joined to the given point by a line that cuts V twice or is tangent to V.

The lines of a three-space R of X determine points of V which lie in a five-space Σ of S. Hence,

Every point of S not on V determines a unique five-space of S.

If $P = (a_1, a_2, \cdots, a_{10})$ is a point of S not on V, the numbers $B_1, B_2, \cdots, B_5$ defined above are not all zeros. We next prove the following theorem:

If P is a point of S not on V, the three-space R in X determined by P is $B_5x_1 - B_4x_2 + B_3x_3 - B_2x_4 + B_1x_5 = 0$.

Let $x = (x_1, \cdots, x_5)$ and $y = (y_1, \cdots, y_5)$ be two points of X that

---

[5] The question as to whether the two non-zero coordinates of P are the same or different is connected in the geometry with the selection of the unit point in X, and in the groups with the selection of a particular element in a cyclic group for a generator.

satisfy the above equation, and let the point of V determined by the line xy be $Q = (b_1, b_2, \cdots, b_{10})$. The condition that $Q + \lambda P$ be on V gives five quadratic equations in λ. Under the conditions on x and y these equations have solutions. Since Q is on V, $\lambda = 0$ is a solution of the five quadratics. Removing the factor λ the five quadratics become the five linear equations

$$\sum b_j \partial B_i / \partial a_j - \lambda B_i = 0, \qquad i = 1, \cdots, 5.$$

The fact that x and y satisfy the equation above gives the relations:

$$B_4 b_1 - B_3 b_2 + B_2 b_3 - B_1 b_4 = 0, \qquad B_5 b_1 - B_3 b_5 + B_2 b_6 - B_1 b_7 = 0,$$
$$B_5 b_2 - B_4 b_5 + B_2 b_8 - B_1 b_9 = 0, \qquad B_5 b_3 - B_4 b_6 + B_3 b_8 - B_1 b_{10} = 0.$$

Making use of these it is easily verified that the λ given by one of the linear equations satisfies the others.

Consider two points P_1 and P_2 in S, neither on V. Each point determines a three-space R_i in X and a five-space Σ_i in S. If R_1 and R_2 are the same, then the line P_1P_2 lies in the five-space Σ_1 $(= \Sigma_2)$. Every point of the line P_1P_2 determines the same three-space R_1 $(= R_2)$ in X; a point of intersection of P_1P_2 with V is the image of a line in R_1. A line, plane, or three-space in S which lies in a five-space Σ determined by the lines of a three-space R in X will be called a *Σ-line*, a *Σ-plane*, or a *Σ-three-space*.

In dealing with Σ-spaces in S we may forget about the four-space X and retain only the three-space R. The five-space Σ intersects V in the four-dimensional hyperquadric determined by the lines of R. A Σ-line may be a ruling, a tangent, a secant, or it may have no point on the hyperquadric.[6] A ruling, a tangent, or a line with two points on V is always a Σ-line.

If the three-spaces R_1 and R_2 determined by P_1 and P_2 are distinct, then P_1P_2 is not a Σ-line. R_1 and R_2 will have a plane σ in common. The lines of σ determine points of a plane π on V, and π is in Σ_1 and Σ_2. Since π and P_1 are both in Σ_1, the polar of P_1 with respect to the intersection of Σ_1 and V intersects π in a line l_1. Likewise, the polar of P_2 with respect to the intersection of Σ_2 and V intersects π in a line l_2. The lines l_1 and l_2 may be distinct, or they may not.

If l_1 and l_2 are distinct, they intersect in a point M which is on V since it is in π. The lines joining M to P_1 and P_2 are both tangent to V at M. Consequently, the plane P_1P_2M is tangent to V at M, and every line in it through M is tangent to V. The point M is uniquely determined by P_1 and

[6] The condition for intersection of the line and V may lead to a quadratic irreducible in GF(p).

P_2; it is the only point of V from which tangents can be drawn to both P_1 and P_2. However, if P_1 and P_2 are moved along the line joining them, M remains unchanged. M is the image of a line m in X; m is in both R_1 and R_2. If P_1 is left fixed and P_2 is moved along the line P_1P_2, R_1 remains fixed but R_2 changes, and σ moves through the pencil of planes in R_1 on m. Thus the line m, and its image M on V, belongs to the line P_1P_2 and not to the pair of points P_1 and P_2.

We consider further the line P_1P_2. R_1 and R_2 are distinct and so are l_1 and l_2. Let Q_1 be a point other than M on l_2. The line P_1Q_1 is in Σ_1; it is not tangent to V at Q_1 since Q_1 is not on l_1. Hence, P_1Q_1 intersects V in a second point Q'_1. Likewise, if Q_2 is selected on l_1 different from M, P_2Q_2 is in Σ_2 and intersects V in a second point Q'_2. The plane $P_1Q_1Q_2$ is in Σ_1 since π is in Σ_1; hence it intersects V in a conic which is degenrate since it contains the line Q_1Q_2 and consists of two intersecting lines since Q'_1 is not on Q_1Q_2. The vertex of the conic is Q_2 since Q_2 is on l_1 and P_1Q_2 is a tangent. The line $Q_2Q'_1$ is therefore a line on V and hence the lines q_2 and q'_1 in X intersect. Similarly, the lines q_1 and q'_2 intersect. Also, the line m intersects both q'_1 and q'_2. The five lines q_1,, q'_1, q_2, q'_2, and m have the following relations:

q_1, q_2, and m form a triangle in σ;
q'_1 intersects q_2 and m, is not in R_2, but is in R_1;
q'_2 intersects q_1 and m, is not in R_1, but is in R_2.

We select a frame of reference in X as follows:

A_1 is the intersection of q_1 and q_2;
A_2 is the intersection of q_1 and m, and is on q'_2;
A_3 is the intersection of q_2 and m, and is on q'_1;

A_4 and A_5 are on q'_1 and q'_2 respectively, and not in σ. With this coordinate system, we have in S

$$Q_1 = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0), \qquad Q_2 = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0),$$
$$Q'_1 = (0, 0, 0, 0, 0, 0, 0, 1, 0, 0), \qquad Q'_2 = (0, 0, 0, 0, 0, 0, 1, 0, 0, 0),$$

$M = (0, 0, 0, 0, 1, 0, 0, 0, 0, 0)$. With a proper choice of the unit point, we have

$$P_1 = (1, 0, 0, 0, 0, 0, 0, 1, 0, 0), \qquad P_2 = (0, 1, 0, 0, 0, 0, 1, 0, 0, 0).$$

We may denote the line P_1P_2 by $k\,l\,0\,0\,0\,0\,l\,k\,0\,0$. We are thus introducing corrdinates k and l on the line, using P_1 and P_2 for vertices of the frame of

reference on the line. An arbitrary point of the line determines the B's: $B_1 = k^2, B_2 = -l^2, B_3 = B_4 = 0, B_5 = kl$. A point of the line on V would require $k^2 = l^2 = 0$.

A second possibility above was that the points P_1 and P_2 were such that R_1 and R_2 were distinct, but such that the lines l_1 and l_2 *in* π coincided. In that case, if M is any point of l_1, then the plane MP_1P_2 is tangent to V at M. In order to investigate this we may take l_1, which is any ruling of V, to consist of the points

$$M_0 = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0), \qquad M_\lambda = (\lambda, 1, 0, 0, 0, 0, 0, 0, 0, 0),$$

$M_\infty = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. The spaces tangent to V are: at M_0, $a_6 = a_7 = a_{10} = 0$; at M_λ, $-a_6 + \lambda a_8 = 0$, $-a_2 + \lambda a_9 = 0$, $a_{10} = 0$; at M_∞, $a_8 = a_9 = a_{10} = 0$. If P_1P_2 is in two of these tangent spaces, then every point of it must have $a_6 = a_7 = a_8 = a_9 = a_{10} = 0$. Such a line is then P_1P_2, where

$$P_1 = (a_1, a_2, a_3, a_4, a_5, 0, 0, 0, 0, 0); \qquad P_2 = (b_1, b_2, b_3, b_4, b_5, 0, 0, 0, 0, 0).$$

The point $kP_1 + lP_2$ is on V for $a_5k + b_5l = 0$. Since P_1P_2 is by hypothesis not a Σ-line it can have no more than one point on V. We may then state the theorem:

The lines of S which are not Σ-lines belong to two types which are distinguished by the properties that a line of one type has no point on V, a line of the other type has one point on V. A line of the first type determines a unique point on V, a line of the second type determines a unique line on V.

The space tangent to V at a point M of V is six-dimensional. It may be verified easily that a Σ-line is always in more than one tangent space, and thus every line of S is in the tangent space at some point of V. The space tangent to V contains planes, three-spaces, etc.; such will be called *τ-planes, τ-three-spaces*, etc. As will appear a little later, τ-planes and Σ-lines will be very useful in distinguishing among space in S. It is clear that the property of being a Σ-line, or a τ-plane, is invariant under collineations of X.

3. Types of points, of lines, and of planes in S.

The points of S.

1. 1 0 0 0 0 0 0 0 0 0, a point of V.
2. 1 0 0 0 0 0 0 1 0 0, a point not on V.

The lines of S.

1. $k\,l\,0\,0\,0\,0\,0\,0\,0\,0$, a ruling of V.
2. $k\,0\,0\,0\,0\,0\,0\,l\,0\,0$, a secant of V.
3. $k\,0\,0\,0\,l\,0\,0\,k\,0\,0$, a tangent to V.
4. $k\,l\,0\,0\,0\,rl\,0\,k\,0\,0$ (r not a square), a Σ-line with no point on V.
5. $k\,0\,0\,l\,0\,0\,0\,k\,0\,0$, not a Σ-line, one point on V.
6. $k\,l\,0\,0\,0\,0\,l\,k\,0\,0$, not a Σ-line, no point on V.

The planes[7] *of S.*

We list first the six types of Σ-plane; all are in the five-space Σ determined by the three-space $x_5 = 0$ in X.

1. $k\,l\,0\,0\,m\,0\,0\,0\,0\,0$, the image of a plane of lines in X. (xvi)
2. $k\,l\,m\,0\,0\,0\,0\,0\,0\,0$, the image of a bundle of lines in X. (xv)
3. $k\,l\,0\,0\,0\,0\,0\,m\,0\,0$, intersecting V in two lines. (xiii)
4. $k\,l\,m\,0\,0\,0\,0\,k\,0\,0$, intersecting V only in the line $k=0$. (xiv)
5. $k\,l\,0\,0\,0\,m\,0\,k\,0\,0$, intersecting V in the conic $k^2 - lm = 0$. (xii)
6. $k\,l\,m\,0\,0\,rl\,0\,k\,0\,0$, ($r$ not a square), intersecting V only in the point $P_3 : k = l = 0$. (xiii)

The rest of the planes belong to no Σ. We group them in sets according to their intersections with V.

Planes with no point on V.

7. $k\,l\,0\,0\,0\,m\,l\,k{+}rm\,m\,0$ ($x^3 + rx - 1$ irreducible), a τ-plane. (v)
8. $k\,l\,m\,0\,{-}rm\,0\,l\,k\,0\,0$ (r not a square), not a τ-plane but contains the Σ-line $l = 0$. (iii)
9. $k\,l\,0\,0\,m\,0\,l\,k\,0\,m$, not a τ-plane and contains no Σ-line. (i)

Planes with at least one point on V, but with no line on V.

10. $k\,l\,0\,0\,m\,0\,l\,k\,0\,0$, tangent to V at P_3, one point on V. (viii)
11. $k\,l\,0\,0\,0\,m\,l\,k\,0\,0$, a τ-plane, one point on V, one line tangent to V. (vii)
12. $k\,l\,m\,0\,0\,0\,l\,k\,0\,0$, not a τ-plane, one point on V, one line tangent to V. (iv)
13. $k\,l\,0\,0\,0\,m\,l{-}rm\,k\,0\,0$ (r not a square), a τ-plane, one point on V, no line tangent to V. (vi)

[7] The roman numbers locate these planes in Todd's list.

14. $k\,l\,m\,rm\,0\,0\,l\,k\,0\,0$, not a τ-plane, one point on V, no line tangent to V. (ii)
15. $k\,l\,0\,0\,0\,0\,l\,k\,0\,m$, not a τ-plane, one point on V, no line tangent to V. (ii)
16. $k+m\,l\,0\,0\,0\,0\,l\,k\,0\,0$, a τ-plane, two points on V. (v)
17. $k+m\,l\,m\,0\,0\,0\,l\,k\,0\,0$, not a τ-plane, two points on V. (iii)
18. $k+m\,l+m\,0\,0\,0\,0\,l\,k\,0\,0$, three points on V. (v)

Planes with a line on V.

19. $k\,0\,0\,0\,0\,0\,0\,0\,l\,m$, a line and an additional point on V. (x)
20. $k\,0\,0\,0\,l\,0\,0\,k\,m\,0$, a line on V and tangent at a point of it. (xi)
21. $k\,0\,0\,0\,0\,0\,l\,k\,m\,0$, a τ-plane, one line on V, not tangent at any point of the line. (ix)
22. $k\,0\,0\,0\,0\,0\,0\,k\,l\,m$, not a τ-plane, one line on V. (ix)

Each of the spaces in the above list is given in terms of a particular coordinate system, but with each is given a set of properties sufficient to differentiate among the spaces and these properties are not dependent on any coordinate system. Planes 14 and 15 provide an exception to the above statement, and we shall determine some geometric properties that will enable us to distinguish between them.

For plane 14, B_1, B_2, B_3, B_4, B_5 become k^2, $-l^2$, $-lm$, rkm, kl, respectively, while for plane 15 they become k^2, $-l^2$, km, lm, kl. In each case the conditions for a point on V are $k=l=0$, and hence each plane intersects V only at the point P_3. Each plane contains the line $P_1P_2 = k\,l\,0\,0\,0\,0\,l\,k\,0\,0$. P_1P_2 is in the space tangent to V at the point $M = 0\,0\,0\,0\,1\,0\,0\,0\,0\,0$; this tangent space is $a_3 = a_4 = a_{10} = 0$ which contains no other point of either plane. Hence, neither plane is a τ-plane. In plane 14 the point P_3 is $0\,0\,1\,r\,0\,0\,0\,0\,0\,0$ and the space tangent to V at P_3 is: $a_5 = 0$, $ra_6 - a_7 = 0$, $ra_8 - a_9 = 0$. The intersection of this tangent space with the plane requires $k = l = 0$, and hence it intersects the plane only at P_3. The space tangent to V at P_3 in plane 15 is $a_1 = a_2 = a_5 = 0$, and it intersects plane 15 only at P_3. So that the only difference between planes 14 and 15 might conceivably be due entirely to the choice of coordinate systems in the two cases. To show that this is not the case, we examine some new geometric relations between S and X.

We recall that if P is a point of S not on V, the numbers $B_1, \cdots, B_5$ are not all zero and that $B_5x_1 - B_4x_2 + B_3x_3 - B_2x_4 + B_1x_5 = 0$ is the three-space R in X determined by P. We now think of the above equation in

another way. Let us suppose that $A = (x_1, x_2, x_3, x_4, x_5)$ is a given point in X. Then the equation becomes a quadratic in the coordinates $a_1, \cdots, a_{10}$ of a point in S. This eight-dimensional hyperquadric in S intersects a plane in S in a conic or else contains the plane completely. Thus every point of X determines in a plane π of S a conic, which may be identically zero. In the cases of planes 14 and 15 the conics given by A are

$$x_1kl - x_2rkm - x_3lm + x_4l^2 + x_5k^2 = 0,$$
$$x_1kl - x_2lm + x_3km + x_4l^2 + x_5k^2 = 0,$$

respectively. There is thus determined in each of the planes a four-parameter set W of conics. The points of a line in X determine a pencil of conics belonging to W. Each of planes 14 and 15 determines a special line in X, namely, the line p_3 which is imaged in S on the point P_3 common to the plane and V. Thus in each case the set W of conics contains a special pencil. For 14 the line p_3 joins A_1 and $A_4 + rA_5$ and the special pencil of conics is $\lambda kl + l^2 + rk^2 = 0$. For 15 the line p_3 joins A_4 and A_5 and the special pencil of conics is $\lambda l^2 + k^2 = 0$. Since $\lambda^2 - 4r = 0$ has no solution in GF(p), every conic in the special pencil in 14 consists of a pair of distinct lines. In the case of 15 $\lambda = 0$ and $\lambda = \infty$ give the conics $k^2 = 0$ and $l^2 = 0$. This proves that no change of coordinates will put 14 into 15.

It is comparatively easy to see that the planes listed above have the properties that are listed with them. This comes from the fact that use has been made of special elements in determining the canonical forms. Plane 9 is identified by its lack of special elements of the types that are used for the others. This plane does, however, contain a special element and the given form depends on proper use being made of it.

In the case of plane 9, the five conics given by the vertices of the frame of reference in X are linearly independent. Five linearly independent conics in a plane determine a unique conic which is apolar to all the conics of the four-parameter set W dependent on them. In plane 9 this conic is $C: m^2 - 2kl = 0$. P_1 and P_2 are points of C and P_3 is the pole of the line P_1P_2 with respect to C. Conversely, if π is any plane of S which has no point on V, which contains no Σ-line, and which is not a π-plane, and if P_1, P_2, and P_3 are selected to have the above relations to C, then a coordinate system in X can be selected so that π has the form 9.

4. Three-spaces in S. We separate the three-spaces of S into sets according to their intersections with V. There is one type of S_3 in S which lies wholly on V; it consists of the points which are images in S of all the

lines in X through a single point. This S_3 is of no use in connection with the groups since it corresponds to the direct product of a cyclic group of order p and a metabelian group generated by four elements. We omit this S_3 from the list.

Three-spaces with at least a non-degenerate conic on V.

1. $k\,0\,m\,0\,n\,0\,0\,l\,0\,0$, the ruled quadric $kl + mn = 0$ on V.
2. $k\,l\,m\,0\,n\,{-l}\,0\,k\,0\,0$, the quadric $k^2 + l^2 + mn = 0$ on V; the quadric has no rulings.
3. $k\,l\,n\,0\,0\,l\,0\,m\,0\,0$, the cone $km - l^2 = 0$ on V.
4. $n\,n\,0\,0\,k\,m\,0\,0\,m\,l$, the conic $kl - m^2 = 0$, $n = 0$, and the line $l = m = 0$ on V.
5. $k\;l{+}n\;m\;0\;n\;n\;l\;k{+}n\;0\;0$, the conic $k^2 + kn - n^2 + mn = 0$, $l = 0$ and the point $(1\,1\,0\,{-1})$ on V.
6. $n\,0\,0\,0\,k\,m\,0\,n\,m\,l$, the conic $kl - m^2 = 0$, $n = 0$, on V.

Three-spaces with a plane on V, not lying wholly on V.

7. $k\,l\,m\,0\,n\,0\,0\,0\,0\,0$.
8. $k\,l\,m\,0\,0\,0\,n\,0\,0\,0$.
9. $k\,l\,m\,0\,n\,0\,0\,0\,0\,n$.
10. $k\,l\,0\,0\,m\,0\,0\,0\,0\,n$.
11. $k\,l\,0\,0\,m\,n\,0\,0\,n\,0$.

The planes on V are of two types: (1) planes whose points represent the lines of a plane in X, and (2) planes whose points represent the lines in a bundle. In spaces 7, 8, 9 above the plane $n = 0$ is of the second type; the first has two planes on V, the second has a plane and a line, and the third has only $n = 0$ on V. Spaces 10 and 11 have no plane of the second type; 11 is in the space tangent to V at P_3, while space 10 is not in the tangent space at any point of V.

Three-spaces with at least two rulings but no plane on V.

12. $k\,l\,0\,0\,0\,m\,0\,0\,0\,n$.
13. $k\,l\,0\,n\,0\,m\,n\,0\,0\,0$.
14. $k\,l\,0\,0\,0\,m\,0\,0\,n\,n$.
15. $k\,l\,0\,n\,0\,m\,0\,n\,0\,0$.
16. $k\,l\,0\,n\,0\,m\,n\,n\,0\,0$.
17. $k\,l\,0\,0\,n\,m\,0\,0\,0\,n$.
18. $k\,l\,0\,n\,n\,m\,0\,0\,0\,0$.

The first two have three rulings of V; in 13 the rulings pass through a point but in 12 they do not. Each of the others contains only two rulings and they intersect; 14 has an additional point on V and the others do not. 15 and 18 are in spaces tangent to V; 16 and 17 are not τ-spaces. 18 is in the space tangent to V at a point of its intersection with V; 15 is not.

To see that 16 and 17 are distinct types, we note that each contains two intersecting rulings of V and coordinate systems have been selected so that in both cases the line A_1A_2, in X, is imaged on this intersection. The line A_1A_2 is therefore special in both cases. An arbitrary point P in 16 determines the three-space R in X:

$$n^2x_1 - n^2x_2 + mnx_3 + lnx_4 + (kn - lm)x_5 = 0.$$

An arbitrary point in 17 gives the three-space

$$n^2x_1 - lnx_2 + knx_3 - lmx_5 = 0.$$

In the case of 16 every point P determines an R which contains the point 1 1 0 0 0, a point of the special line; in the case of 17 no point of the special line is in every R.

Three-spaces containing one ruling of V.

19. $m\ 0\ 0\ k\ 0\ 0\ l\ n\ k\ l$.
20. $k\ l\ 0\ 0\ 0\ n\ 0\ 0\ n\ m$.
21. $k\ l\ m\ n\ 0\ rn\ m\ 0\ 0\ 0$.
22. $k\ l\ 0\ 0\ 0\ n\ m\ m\ rn\ 0$.
23. $k\ l\ m\ 0\ m\ 0\ n\ n\ 0\ 0$.
24. $k\ l\ 0\ 0\ n\ 0\ m\ m\ 0\ -rn$.
25. $k\ l\ m\ 0\ n\ 0\ 0\ 0\ m\ n$.
26. $k\ l\ n\ 0\ 0\ 0\ m\ m\ n\ 0$.
27. $k\ l\ m\ n\ 0\ 0\ m\ n\ 0\ 0$.

In 21., 22., and 24., r is not a square.

Spaces 19 and 20 have respectively two and one points on V in addition to the ruling $m = n = 0$; the rest have only the ruling. Space 21 is in the space tangent to V at P_1; 22 is in the space tangent to V at a point not in 22; none of the others is a τ-space. In 23 every plane on P_1 is a τ-plane and every line in the plane $n = 0$ is tangent to V at its intersection with P_1P_2. Space 24 contains the τ-plane $n = 0$ and it contains a pencil of τ-planes on P_3P_4. P_3P_4 is a Σ-line which does not intersect the ruling. All τ-planes pass through P_3 but not every plane on P_3 is a τ-plane. Space 25 contains a pencil of τ-planes and no others; the axis of the pencil is P_2P_3, which is not the ruling. Space 26 contains a pencil of τ-planes on the ruling P_1P_2; one of the τ-planes is tangent to V at P_2. Space 27 contains a pencil of τ-planes on P_1P_2; one plane is tangent at P_1, another is tangent at P_2.

Three-spaces interesecting V in three or more points, but not in any plane curve.

28. $k{+}n\ k\ 0\ 0\ 0\ n\ l\ m\ n\ 0$.
29. $k\ k\ n\ -n\ -n\ 0\ l\ m\ 0\ n$.
30. $k\ k\ n\ -n\ 0\ 0\ l\ m\ 0\ n$.
31. $k\ k\ n\ n\ 0\ 0\ l\ m\ n\ n\ n$.
32. $k\ k\ n\ -n\ n\ 0\ l\ m\ 0\ n$.
33. $k\ k\ 0\ n\ n\ 0\ l\ m\ 0\ n$.
34. $k\ k\ 0\ n\ 0\ n\ l\ m\ 0\ 0$.
35. $k\ k\ n\ n\ n\ n\ l\ m\ 0\ n$.

The space 28 intersects V in a twisted cubic curve; 29 intersects V in five points; 30 intersects V in four points and contains a Σ-line tangent to V at one of them; 31 intersects V in four points but contains no tangent line at any of them; each of the others has just three points on V. Space 32 contains a Σ-line not in $n = 0$ and not tangent to V; 33 contains one Σ-line tangent to V; 34 contains two Σ-lines tangent to V at two distinct points; 35 contains no Σ-line except the sides of the triangle determined by the three points of V.

Three-spaces with two points on V.

36. $k+n\ l\ n\ 0\ 0\ m\ l\ k+rm\ m\ 0$, $x^3 + rx - 1$ irreducible.
37. $k\ l\ 0\ -n\ n\ 0\ l\ m\ 0\ n$.
38. $k\ l\ 0\ 0\ n\ 0\ l\ m\ 0\ n$.
39. $k\ l\ 0\ n\ rn\ 0\ l\ m\ 0\ 0$.
40. $k\ l\ 0\ n\ 0\ n\ l\ m\ n\ 0$.
41. $k\ l\ 0\ n\ 0\ n\ l\ m\ 0\ 0$.
42. $k\ l\ n\ n\ n\ 0\ l\ m\ 0\ 0$.

Space 36 contains the τ-plane $n = 0$ which has no point on V; the others contain τ-planes every one of which has at least one point on V. Spaces 37 and 38 have three τ-planes each; in 37 one of the τ-planes contains both points of V, in 38 two of the τ-planes contain both points of V. All of the planes on P_3 in 39 are τ-planes, and so is the plane $m = 0$. Space 40 contains two τ-planes. Spaces 41 and 42 have pencils of τ-planes on the line joining the two points of V and in each the plane $m = 0$ is a τ-plane. To distinguish between 41 and 42 an argument analogous to that given for planes 14 and 15 suffices. The points of X determine in each of spaces 41 and 42 a four-parameter system W of quadrics. In W there are two special pencils of quadrics determined by the two points on V. In 41 each quadric in either of the special pencils consists of a pair of planes; in 42 one of the pencils contains two quadrics each of which is a plane counted twice.

Three-spaces with one point on V.

43. $k\ l\ 0\ 0\ n\ m\ l\ k+rm\ m\ 0$, $x^3 + rx - 1$ irreducible.
44. $k\ l\ 0\ n\ m\ n\ l\ k\ 0\ 0$.
45. $k\ l\ n\ 0\ n\ m\ l\ k\ 0\ 0$.
46. $k\ l\ 0\ -n\ n\ m\ l\ k\ 0\ 0$.
47. $k\ l\ n\ 0\ 0\ m\ l\ k\ n\ 0$.
48. $k+n\ l\ m\ 0\ 0\ rn\ l\ k\ n\ 0$, $x^3 - x - r^2$ irreducible.
49. $k\ l\ n\ 0\ 0\ m\ l+m\ k\ n\ 0$.
50. $k\ l\ 0\ -n\ n\ m\ l+m\ k\ 0\ 0$.
51. $k\ l\ n\ -n\ n\ 0\ l\ k\ 0\ m$.
52. $k\ l\ n\ n\ -n\ 2n\ l\ k\ 0\ m$.

The space 43 is tangent to V at P_4 which is on V; none of the others has this property. Spaces 44 and 45 each contain a plane tangent to V at O, the point on V; in 45 this tangent plane is a Σ-plane, in 44 it is not.

Spaces 46, 47, 48 intersect the space tangent to V at O in a line; 46 contains two τ-planes, 47 and 48 each contain only one τ-plane; in 47 the τ-plane passes through O, in 48 it does not. The space tangent to V at O intersects none of the other spaces anywhere except at O. Space 49 contains a single τ-plane, space 50 contains two τ-planes. Spaces 51 and 52 contain no τ-planes; each contains a special pencil of quadrics belonging to the four-parameter set W determined by the points of X. In both cases the quadrics of the special pencils are cones; in 51 three of the cones are degenerate, each consists of a pair of planes, in 52 only one is degenerate.

Three-spaces with no point on V.

53. $k\,l\,m\,0\,m\,n\,l\,k{+}n\,n\,0$. 54. $k\,l\,0\,2n\,m{+}3n\,n\,l\,k\,0\,m$.

Space 53 contains a τ-plane and a Σ-line; 54 contains neither. In 53 the τ-plane is $m = 0$ and the Σ-line $l = n = 0$. These special elements are made use of to get the canoncial form. There are no special lines or planes in 54. These facts are easily verified and are enough to distinguish between 53 and 54.

We mention another way in which 53 and 54 can be distinguished. In 54 the plane $n = 0$ is number 9 of the list of planes; every plane of 54 is of this type. Every plane of 53 except the τ-plane and the planes of the pencil on the Σ-line is of the same type. In either case we may take the plane $P_1P_2P_3$ to be $k\,l\,0\,0\,m\,0\,l\,k\,0\,m$ and $P_4 = 0\,0\,a_3\,a_4\,a_5\,a_6\,a_7\,a_8\,a_9\,0$. In either case the conditions for a single point on V lead to a polynomial of degree five in a single variable. In case 54 the polynomial is irreducible; in case 53 the polynomial is the product of a quadratic and a cubic, both irreducible in GF(p).

5. Some remarks about the groups. We make a few observations about the groups determined in the preceding pages. There are 85 groups: 1 of order p^{15}; 2 of order p^{14}; 6 of order p^{13}; 22 of order p^{12}; 54 of order p^{11}. Two groups of the same order have commutator subgroups of the same order; the index of the commutator subgroup is always p^5.

Let us change the notation to denote the commutators of pairs of generators by s_k instead of c_{ij} so that the commutator of x and y is $\Pi s_k^{a_k}$, where $1 \leqq k \leqq 10$. The group G, of order p^{15}, was defined by relations on the generators of which one was the requirement that there be no relations among the s_i's. For each of the other groups there will be relations among them.

Let S_i, $i = 0, 1, 2, 3$, be a subspace of dimension i of S. Let H be the

subgroup of order p^{i+1} of C which is determined by S_i. Let G' and C' be the quotient groups G/H and C/H respectively. Then G' will be defined by the relations which define G together with $i+1$ relations which put equal to the identity commutators which correspond to $i+1$ independent points of S_i; ordinarily the simplest relations are obtained if commutators corresponding to the vertices of the frame of reference in S_i are put equal to the identity. For example, if S_2 is plane 9, the simplest additional relations required to define G' are: $s_1s_8=1$, $s_2s_7=1$, $s_5s_{10}=1$.

We have given close attention to the subgroups of G determined by subspaces of S; it will be useful to look more closely at some subgroups of G determined by subspaces of X. A point in X determines a subgroup α of order p in the abelian quotient group G/C and determines an abelian subgroup of order p^{11} in G. Two abelian subgroups of order p^{11} of G are conjugate under the group of isomorphisms of G, but every abelian subgroup of order p^{11} is self-conjugate in G. A line in X determines a subgroup of order p^{12} of G and this subgroup is non-abelian with commutator subgroup of order p. This subgroup is also self-conjugate in G but is conjugate to any other such subgroup under the group of isomorphism of G. Such a subgroup is $\{C, U_1, U_2\}=\Gamma$. Γ is the direct product of the metabelian group $\{U_1, U_2\}$ which contains the commutator s_1 and the abelian group $\{s_2, s_3, \cdots, s_{10}\}$. Γ is also the direct product of the abelian group just given and any group $\{cU_1, c'U_2\}$ where c and c' are arbitrary elements of C. The group $\{s_2, s_3, \cdots, s_{10}\}$ is not, however, uniquely determined by the line in X since it may be replaced by any subgroup of order p^9 of C which does not contain s_1. A line in X therefore determines a unique subgroup of G of order p^{12}, and determines a set of metabelian subgroups of order p^3 all of which have for a commutator subgroup the cyclic group determined by the point in S which is the image of the line in X. A plane in X determines a subgroup of order p^{13} of G, and a set of metabelian subgroups of order p^6 all of which have the same commutator subgroup of order p^3. A three-space in X determines a unique subgroup of order p^{14}, and a set of metabelian subgroups each generated by four elements, each of order p^{10}, and all with the same commutator subgroup of order p^6.

The groups G' of order p^{14}, determined by the two types of point in S, differ in the fact that the first one contains an abelian subgroup $\{C', U_1, U_2\}$ of order p^{11} and the second contains no abelian subgroup of order p^{11}. The fact that the point of the second type determines a unique three-space R in X is reflected in the fact that the corresponding G' contains a unique subgroup of order p^{13} which has a commutator subgroup of order p^5.

The lines of S determines six types of group of order p^{13}, each with commutator subgroup of order p^8. In G' which corresponds to line 1 of the list of lines there are $p+1$ abelian subgroups $\{C', U_1, U_2^k, U_3^l\}$ of order p^{10} all contained in the non-abelian subgroup $\{C', U_1, U_2, U_3\}$ of order p^{11}. The second line of the list gives a group G' which contains two abelian subgroups of order p^{10}, viz., $\{C', U_1, U_2\}$ and $\{C', U_3, U_4\}$; both abelian subgroups are contained in the group $\{C', U_1, U_2, U_3, U_4\}$ of order p^{12} whose commutator subgroup has order p^4; this last is the only subgroup of order p^{12} with commutator subgroup of order less than p^5. Lines 3 and 4 of the list determine groups G' which have respectively one and no abelian subgroup of order p^{10}. Both have subgroups of order p^{12} with commutator subgroups of order p^4. Lines 5 and 6 determine groups G' which also have respectively one and no abelian subgroups of order p^{10}. Neither contains a subgroup of order p^{12} with commutator subgroup of order as small as p^4.

The group G' determined by line number 6 in the list is defined by the (additional) relations: $s_1s_8 = 1$, $s_2s_7 = 1$. Let us denote by L the line $k\,l\,0\,0\,0\,0\,l\,k\,0\,0$. Every point of L determines a subgroup Γ of G with commutator subgroup β; Γ is of order p^{14} and β is of order p^6. Corresponding to these groups are subgroups Γ' and β' of order p^{12} and p^5 in G'. The line L determines the point $M = 0\,0\,0\,0\,1\,0\,0\,0\,0\,0$ on V. The line m is in the three-space R determined in X by any point P of L. Hence, the non-abelian group $\{C', U_2, U_3\}$ of order p^{10} is in every one of the $p+1$ special subgroups Γ' of G', which are special because they are the only subgroups of order p^{12} with commutator subgroups of order p^5.

The above examples are meant to suggest the complexity of the relations among subgroups of the groups G', but also to suggest that these complex relations are not, in these groups at least, unmanageable.

UNIVERSITY OF ILLINOIS.

HYPERGEODESICS AND DUAL HYPERGEODESICS.*

By T. K. Pan.

1. Introduction. Dual union curves on a surface were first defined and studied by P. Sperry [4, p. 222] and investigated later by G. M. Green [2, pp. 140-144]. In this paper a system of curves on a surface called dual hypergeodesics is defined, which includes a system of dual union curves as a special case, its intrinsic differential equation is found, most of the theorems for dual union curves are generalized and some properties of hypergeodesics are studied.

The notation of Eisenhart [1] will be used for the most part except that $\Gamma^{\alpha}{}_{\beta\gamma}$ will be employed for Christoffel symbols of the second kind.

2. The extended relation *R*. Consider a real proper analytic surface S defined with reference to a rectangular cartesian coordinate system by

$$x^i = x^i(u^1, u^2) \qquad (i = 1, 2, 3). \tag{2.1}$$

The coordinate curves are assumed to form the asymptotic net. The tangents to one family of the net constructed at the points of a curve of the other family generate a non-developable ruled surface. There are two such surfaces R_1, R_2 associated with $P(x^i)$. Any two planes different from the tangent plane of S at P and passing through the asymptotic tangents at P will be tangent to R_1, R_2 at two points P_1, P_2 respectively. The intersecting line L' of these two planes passes through P but does not lie on the tangent plane of S at P; and the line joining P_1, P_2 lies on the tangent plane of S at P but does not pass through P. These two lines L', L are said to be in the relation R with respect to the asymptotic net [3, p. 81]. As P varies over S, they generate two congruences Γ', Γ respectively.

Let the direction cosines of L' be defined by

$$\lambda^i = p^{\alpha}\partial x^i/\partial u^{\alpha} + qX^i \tag{2.2}$$

where X^i are the direction cosines of the normal to S at P, and where p^{α}, q

* Received May 16, 1950.

are functions of u^α, q being assumed to be always positive. The coordinates of the two points P_α are found to be

$$\bar{x}^i{}_\alpha = x^i - q/(q\Gamma^\beta{}_{12} - d_{12}p^\beta)\partial x^i/\partial u^\alpha, \qquad \alpha \neq \beta. \tag{2.3}$$

In terms of the local coordinates y_1, y_2 which are defined as the coordinates of the terminal point of the vector

$$y_1\left(\frac{1}{\sqrt{g_{11}}}\frac{\partial x^i}{\partial u^1}\right) + y_2\left(\frac{1}{\sqrt{g_{22}}}\frac{\partial x^i}{\partial u^2}\right)$$

when it is localized at P, the equation of L is readily found to be

$$\sqrt{g_{22}}(q\Gamma^2{}_{12} - d_{12}p^2)y_1 + \sqrt{g_{11}}(q\Gamma^1{}_{12} - d_{12}p^1)y_2 + q\sqrt{(g_{11}g_{22})} = 0. \tag{2.4}$$

When the asymptotic tangents at P are considered as the lines L', the lines L in the relation R to them with respect to the asymptotic net are defined to be given by (2.4). We call the relation R with the addition of these pairs of lines *the extended relation* R. Geometrically, an asymptotic tangent at P on S is in the extended relation R to itself with respect to the asymptotic net.

3. Definition and equation of dual hypergeodesics. The integral curves of a differential equation of the form

$$d^2u^2/du^{1^2} = A + Bdu^2/du^1 + C(du^2/du^1)^2 + D(du^2/du^1)^3 \tag{3.1}$$

in which A, B, C, D are analytic functions of u^α, are called hypergeodesics on S. They form a two-parameter family. The envelope of their osculating planes through P on S is a cone with vertex at P and has the tangent plane of S at P for bitangent plane, the lines of contact being the asymptotic tangents of S at P. This cone which is called the osc-cone at P of the family is ordinarily of class three and order four [3, p. 101]. Except the two asymptotic tangents all generators of the osc-cone at P are lines L'. Hence we can always find the line L which is in the extended relation R to each generator of the osc-cone at P of a family of hypergeodesics.

A dual hypergeodesic relative to a family of hypergeodesics will be defined as a curve on S such that the ray point at each point P of the curve lies on the envelope of the lines L which are in the extended relation R to the generators of the osc-cone at P of the family with respect to the asymptotic net of curves on S.

Let $u^2 = u^2(u^1)$ be a hypergeodesic C through P of the family (3.1)

such that its direction at P is given by $\lambda = du^2/du^1$. The osculating plane of C at $P(x^i)$ is found by the aid of the equations of Gauss to be

$$(3.2) \quad e_{ijk}(\bar{x}^i - x^i)\left(\frac{\partial x^j}{\partial u^1} + \frac{\partial x^j}{\partial u^2}\lambda\right)\Big[\frac{\partial x^k}{\partial u^2}(A + B\lambda + C\lambda^2 + D\lambda^3) + \frac{\partial x^k}{\partial u^\gamma}(\Gamma^\gamma{}_{11} + 2\Gamma^\gamma{}_{12}\lambda + \Gamma^\gamma{}_{22}\lambda^2) + 2d_{12}\lambda X^k\Big] = 0.$$

Equation (3. 2) and the equation

$$(3.3) \quad e_{ijk}(\bar{x}^i - x^i)\Big[\frac{\partial x^j}{\partial u^1}\frac{\partial x^k}{\partial u^2}(B + 2C\lambda + 3D\lambda^2) + 2\frac{\partial x^j}{\partial u^1}\frac{\partial x^k}{\partial u^\gamma}(\Gamma^\gamma{}_{12} + \Gamma^\gamma{}_{22}\lambda) + 2d_{12}X^k\frac{\partial x^j}{\partial u^1} + \frac{\partial x^j}{\partial u^2}\frac{\partial x^k}{\partial u^2}(A + 2B\lambda + 3C\lambda^2 + 4D\lambda^3) + \frac{\partial x^j}{\partial u^2}\frac{\partial x^k}{\partial u^\gamma}(\Gamma^\gamma{}_{11} + 4\Gamma^\gamma{}_{12}\lambda + 3\Gamma^\gamma{}_{22}\lambda^2) + 4d_{12}\lambda X^k\frac{\partial x^j}{\partial u^2}\Big] = 0$$

define two planes intersecting in a generator L' of the osc-cone of the family (3. 1) corresponding to the hypergeodesics C in the direction λ at P. The direction numbers of the normal to each of these planes are

$$(3.4) \quad a_m = \sqrt{g}X^m[(A + \Gamma^2{}_{11}) + (B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})\lambda + (C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})\lambda^2 + (D - \Gamma^1{}_{22})\lambda^3] + 2d_{12}e_{mjk}\frac{\partial x^j}{\partial u^1}X^k\lambda + 2d_{12}e_{mjk}\frac{\partial x^j}{\partial u^2}X^k\lambda^2$$

$$(3.5) \quad b_n = \sqrt{g}X^n[(B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11}) + 2(C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})\lambda + 3(D - \Gamma^1{}_{22})\lambda^2] + 2d_{12}e_{njk}\frac{\partial x^j}{\partial u^1}X^k + 4d_{12}e_{njk}\frac{\partial x^j}{\partial u^2}X^k\lambda.$$

Hence the generator L' has the direction numbers

$$(3.6) \quad c^g = e^{gmn}a_m b_n$$

which, by the aid of the formula [1, p. 8, Ex. 9]

$$(3.7) \quad e^{ijh}e_{klh} = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}$$

and after discarding of irrelevant non-zero multipliers, reduces to

$$(3.8) \quad c^i = P^\alpha \partial x^i/\partial u^\alpha + QX^i$$

where

$$
\begin{aligned}
P^1 &= (A + \Gamma^2{}_{11}) - (C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})\lambda^2 - 2(D - \Gamma^1{}_{22})\lambda^3, \\
P^2 &= 2(A + \Gamma^2{}_{11})\lambda + (B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})\lambda^2 - (D - \Gamma^1{}_{22})\lambda^4, \qquad (3.9)\\
Q &= 2d_{12}\lambda^2.
\end{aligned}
$$

Substituting (3.9) into (2.4) and arranging the terms as a polynomial in λ we obtain the following equation of L, which is in the extended relation R to the generator L'

$$a\lambda^4 + b\lambda^3 + c\lambda^2 + d\lambda + e = 0, \tag{3.10}$$

where the coefficients are defined by

$$
\begin{aligned}
&a = \sqrt{g_{22}}(D - \Gamma^1{}_{22})y_1, \qquad b = 2\sqrt{g_{11}}(D - \Gamma^1{}_{22})y_2, \\
&c = -\sqrt{g_{22}}(B - \Gamma^1{}_{11})y_1 + \sqrt{g_{11}}(C + \Gamma^2{}_{22})y_2 + 2\sqrt{(g_{11}g_{22})}, \qquad (3.11)\\
&d = -2\sqrt{g_{22}}(A + \Gamma^2{}_{11})y_1, \qquad e = -\sqrt{g_{11}}(A + \Gamma^2{}_{11})y_2.
\end{aligned}
$$

As λ varies, the line L (3.10) generates an envelope in the tangent plane of S at P, whose equation is found after considerable simplification to be

$$ad^2 + eb^2 + bcd = 0 \tag{3.12}$$

which from (3.11) becomes

$$
\begin{aligned}
&(A + \Gamma^2{}_{11})(y_1/\sqrt{g_{11}})^3 + (-D + \Gamma^1{}_{22})(y_2/\sqrt{g_{22}})^3 \qquad (3.13)\\
&\quad - y_1y_2/\sqrt{(g_{11}g_{22})}[(-B + \Gamma^1{}_{11})y_1/\sqrt{g_{11}} + (C + \Gamma^2{}_{22})y_2\sqrt{g_{22}} + 2] = 0.
\end{aligned}
$$

This envelope has a node at P, the nodal tangents being the asymptotic tangents of the surface at P. The curve is ordinarily of the third order and will be called *the reciprocal cubic* at P of the family (3.1).

The ray point at P of a curve $u^a = u^a(s)$ on S is by definition the intersection of the planes

$$
\begin{aligned}
&\Sigma(\bar{x}^i - x^i)X^i = 0, \qquad \Sigma(\bar{x}^i - x^i)u'^a\partial X^i/\partial u^a = 0, \\
&\Sigma(\bar{x}^i - x^i)\left(\frac{\partial^2 X^i}{\partial u^a \partial u^\beta}u'^a u'^\beta + \frac{\partial X^i}{\partial u^a}u''^a\right) = \frac{\partial X^i}{\partial u^a}\frac{\partial x^i}{\partial u^\beta}u'^a u'^\beta. \qquad (3.14)
\end{aligned}
$$

Solving simultaneously, C. E. Springer found its local coordinates [5, p. 904] to be

$$y_\delta = \sqrt{g_{\delta\delta}} \cdot dd_{a\beta}u'^a u'^\beta \frac{e_{\gamma\sigma}d^{\sigma\delta}u'^\gamma}{e^{\sigma\tau}(d_{\gamma\sigma}d_{\mu\tau}\bar{\Gamma}^\mu{}_{a\beta}u'^a u'^\beta u'^\gamma + d_{a\sigma}d_{\beta\tau}u'^a u''^\beta)} \tag{3.15}$$

where δ is not summed and where $\bar{\Gamma}^\mu{}_{a\beta} = d^{\nu\gamma}\partial d_{a\nu}/\partial u^\beta - d^{\nu\gamma}d_{a\mu}\Gamma^\mu{}_{\nu\beta}$. Sub-

stituting (3.15) into (3.13), we find that the dual hypergeodesics relative to the family (3.1) with respect to the asymptotic net are defined by the equation

$$(3.16) \quad d^2u^2/du^{12} = -A + Bdu^2/du^1 + C(du^2/du^1)^2 - D(du^2/du^1)^3.$$

When (3.1) is a family of union curves, (3.16) evidently reduces to the equation of dual union curves which include dual geodesics [5, p. 904] as a special case.

4. Reciprocal characteristics. Since equation (3.16) is of the same form as (3.1), it defines a two parameter family of hypergeodesics. When (3.1) is given, (3.16) is uniquely determined. Hence

THEOREM 4.1. *The dual hypergeodesics relative to a given two-parameter family of hypergeodesics on a surface are hypergeodesics. They form a two parameter family called the dual of the given family. There exists one and only one family of hypergeodesics which is the dual of a given two-parameter family of hypergeodesics.*

Equation (3.16) differs from equation (3.1) in having the signs of A and D changed. It is evident that each represents the dual of the family of hypergeodesics defined by the other. Hence

THEOREM 4.2. *The dual relation between any two two-parameter families of hypergeodesics on a surface is reciprocal.*

A family of hypergeodesics and the dual of the family will be called *reciprocal families.* Thus (3.1) and (3.16) represent reciprocal families of hypergeodesics. In particular the union curves of a congruence Γ' on a surface are the dual hypergeodesics relative to the family of dual union curves of the congruence Γ, and therefore the geodesics on a surface are the dual hypergeodesics relative to the family of dual geodesics.

An asymptotic line on a surface is always a hypergeodesic on the surface, since the coefficients in (3.1) are arbitrary and $A = 0$ gives $u^2 =$ const. as a solution of (3.1). But from (3.16) $A = 0$ is also the condition that $u^2 =$ const. is a dual hypergeodesic. Hence an asymptotic line is at the same time a hypergeodesic of two reciprocal families of hypergeodesics.

Any straight line on a surface is an asymptotic line and also a union curve on the surface, since a geodesic is a special union curve. The osculating planes of a curved asymptotic line on a surface are tangent to the surface along the curve and consequently can not pass through any line L' of a

congruence Γ'. Hence there is no curved asymptotic line on a surface, which is a union curve on the surface. Thus straight lines on a surface are the only union curves which are asymptotic lines.

If an asymptotic line is a hypergeodesic of every two-parameter family of hypergeodesics (3. 1) for which $A = 0$ or $D = 0$, then it can be a dual union curve and in that case it is a straight line. Hence

THEOREM 4. 3. *If an asymptotic line is a hypergeodesic of every family* (3. 1) *for which $A = 0$ or $D = 0$, it is a straight line.*

When a family of hypergeodesics is identical with its reciprocal, we say the family is *self dual.* Equations (3. 1), (3. 16) become identical if and only if

$$A \equiv 0, \qquad D \equiv 0, \tag{4.1}$$

which by Theorem 4. 3 imply that the asymptotic lines on the surface are straight lines. Consequently the surface is a quadric. Hence

THEOREM 4. 4. *Every two-parameter family of hypergeodesics on a surface is self dual if and only if the surface is a quadric.*

When a two-parameter family of hypergeodesics is not self dual, there may be a particular hypergeodesic of the family coincident with a hypergeodesic of its reciprocal family. From equations (3. 1), (3. 16) such a curve is a solution of the two equations

$$A + D(du^2/du^1)^3 = 0, \tag{4.2}$$

$$d^2u^2/du^{12} = Bdu^2/du^1 + C(du^2/du^1)^2. \tag{4.3}$$

The curves defined by (4. 2), their directions and the tangents in these directions will be called respectively *the generalized curves of Segre, the gneralized directions of Segre,* and *the generalized tangents of Segre, associated with the family* (3. 1) *or* (3. 16).

The solutions of equation (4. 2) do not necessarily satisfy equation (4. 3). There may exist no solution or one solution of (4. 2) which satisfies (4. 3). In case $AD \neq 0$, when two solutions of (4. 2) satisfy (4. 3), it can be shown easily that the third solution will also satisfy it. Hence

THEOREM 4. 5. *The hypergeodesics common to two reciprocal families of hypergeodesics are the generalized curves of Segre associated with either one of these families. Conversely, two reciprocal families of hypergeodesics do not necessarily have hypergeodesics in common except reciprocal families on a quadric surface or those families on a general surface for which*

$A = D = 0$, which are always identical. On a non-ruled surface those reciprocal families for which $AD \neq 0$, may have no hypergeodesics in common or may have one or three one-parameter families of hypergeodesics in common.

5. Conjugate nets. If the family (3.1) contains a conjugate net represented by

$$(du^2)^2 - \lambda^2 (du^1)^2 = 0 \tag{5.1}$$

where λ is a scalar function of u^α, we have

$$\phi_{u^2} = 2A + 2C\phi, \qquad \phi_{u^1} = 2B\phi + 2D\phi^2, \tag{5.2}$$

where A, B, C, D are supposed to be known functions and $\phi = \lambda^2$ is to be determined. A necessary and sufficient condition for a common solution of (5.2) is that

$$\partial(A + C\phi)/\partial u^1 = \partial(B\phi + D\phi^2)/\partial u^2 \tag{5.3}$$

be satisfied identically by such a solution. This condition after differentiation becomes

$$(D_{u^2} + 2CD)\phi^2 + (B_{u^2} - C_{u^1} + 4AD)\phi + (2AB - A_{u^1}) = 0 \tag{5.4}$$

which is either satisfied identically in ϕ at each point or not. In any case we can show that A, B, C, D can not all be arbitrary and obtain the following result.

THEOREM 5.1. *A two-parameter family of hypergeodesics on a surface does not in general contain a conjugate net, but a family can be chosen in infinitely many ways so that its hypergeodesics contain one conjugate net or a one-parameter family of conjugate nets.*

A similar result can be obtained for the pairing of the hypergeodesics of a family (3.1) and the dual hypergeodesics of its reciprocal family (3.16) to form a conjugate net.

The condition that (3.16) will contain the associate conjugate net of the net (5.1), defined by

$$(du^2)^2 + \lambda^2 (du^1)^2 = 0, \tag{5.5}$$

is found exactly the same as (5.2). Hence

THEOREM 5.2. *If there exists on a surface a conjugate net formed of hypergeodesics of a two-parameter family of hypergeodesics, then the associate conjugate net of it consists of the dual hypergeodesics of the reciprocal family; and conversely.*

6. Associated cones and plane curves. In the homogeneus local cartesian coordinates v_1, v_2, v_3, v_4 of the plane $v_1y_1 + v_2y_2 + v_3y_3 + v_4 = 0$, the equation of the cusp axis of the osc-cone of (3.1) at P are found to be

$$(6.1)\quad -\sqrt{g_{11}}(C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})v_1 + \sqrt{g_{22}}(B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})v_2 + 2d_{12}v_3 = 0, \qquad v_4 = 0,$$

and its three cuspidal tangent planes are found to intersect the tangent plane of S at P in the three tangents

$$(6.2)\quad (\sqrt{g_{11}})^3(-D + \Gamma^1{}_{22})v^3{}_1 + (\sqrt{g_{22}})^3(A + \Gamma^2{}_{11})v^3{}_2 = 0, \qquad v_4 = 0.$$

These three tangents will be the same as those determined by the reciprocal family (3.16) if and only if

$$(6.3)\quad A = \rho\Gamma^2{}_{11}, \qquad D = -\rho\Gamma^1{}_{22},$$

where $\rho \neq 1$ is either a constant or a function of u^α. In this case the directions of the three tangents become those of Segre, since $y_1/\sqrt{g_{11}} = dw^1$, $y_2/\sqrt{g_{22}} = dw^2$. Hence

THEOREM 6.1. *The osc-cones of two reciprocal families of hypergeodesics at a point P always have a common cusp axis. They have the same cuspidal tangent planes if and only if the latter intersect the tangent plane of S at P in its Segre tangents.*

The locus of the ray points of all the hypergeodesics through P of a family (3.1) is a plane cubic curve called the ray-point cubic at P of the family. Its equation in local coordinates on the tangent plane is found to be

$$(6.4)\quad (-A + \Gamma^2{}_{11})(y_1/\sqrt{g_{11}})^3 + (D + \Gamma^1{}_{22})(y_2/\sqrt{g_{22}})^3 - y_1y_2/\sqrt{(g_{11}g_{22})}[(-B + \Gamma^1{}_{11})y_1/\sqrt{g_{11}} + (C + \Gamma^2{}_{22})y_2/\sqrt{g_{22}} + 2] = 0.$$

A comparison of (6.4) and (3.13) and some calculation yield the results:

THEOREM 6.2. *The reciprocal cubic of a family of hypergeodesics is the ray-point cubic of its reciprocal family. The ray-point cubics or the reciprocal cubics at a point of two reciprocal families of hypergeodesics have a common flex ray. They possess a common set of points of inflection if and only if the three points are on the Darboux tangents.*

THEOREM 6.3. *The osc-cones of two reciprocal families of hypergeodesics possess three common cuspidal tangent planes if and only if their ray-point cubics or reciprocal cubics possess three common points of inflection.*

Since two reciprocal families have the same cusp axis, they have the same cusp axis curves, which form a conjugate net if and only if $B_{u^2} + C_{u^1} = 0$. Since a self dual family consists of a pencil of conjugate nets if and only if $B_{u^2} - C_{u^1} = 0$, we have

THEOREM 6.4. *The cusp axis curves of a self dual family of hypergeodesics which contains a pencil of conjugate nets form a conjugate net if and only if B is a function of u^1 alone and C is a function of u^2 alone.*

In general the osc-cones at a point of two reciprocal families of hypergeodesics have sixteen generators in common, four of which coincide with the two asymptotic tangents at the point. The remaining twelve common generators are in general distinct. For if two of the non-asymptotic common generators coincide with L', the two osc-cones are tangent to the same plane along L' and this requires $A - 2D\lambda^3 = 0$, $2A - D\lambda^3 = 0$, that is $A = 0$, $D = 0$, since $\lambda \neq 0$ or ∞. Hence

THEOREM 6.5. *The osc-cones of two reciprocal families do not have double contact along any non-asymptotic common generator unless the family is self dual or the surface is a quadric.*

The direction λ at which the osc-cones of two reciprocal families have a common tangent plane along different generators is found to satisfy

$$A + D\lambda^3 = 0. \tag{6.5}$$

Since λ is non-asymptotic, $A = 0$ and $D \neq 0$ or $A \neq 0$ and $D = 0$ can not satisfy (6.5). In case $A = D = 0$, the osc-cones are identical. Hence

THEOREM 6.6. *On a non-ruled surface the osc-cones at P of two reciprocal families for which $AD \neq 0$ have, besides the plane tangent to S at P, three common tangent planes in the generalized directions of Segre associated with either of the families.*

The osc-cones at P of two reciprocal families for which $A = 0$, $D \neq 0$ or $A \neq 0$, $D = 0$ on a non-ruled surface and those for which $A \neq 0$ or $D \neq 0$ on a ruled surface have only one common tangent plane—the tangent plane to the surface at P. This plane corresponds to an asymptotic direction which is now a generalized direction of Segre associated with the respective family under consideration. The osc-cones at P of two reciprocal families for which $A = 0$, $D = 0$ on a non-ruled surface and those on a quadric are identical and have therefore infinitely many common tangent planes corresponding to the infinitely many directions of Segre associated with the

respective family under consideration. With Theorem 6.6 we may sum up these results in the following theorem:

THEOREM 6.7. *The generalized directions of Segre associated with a family of hypergeodesics are the directions in which the osc-cones of the family and its reciprocal family have common tangent planes.*

The following two theorems can be easily verified:

THEOREM 6.8. *The axis cones at a point of two reciprocal families of hypergeodesics have double contact along the asymptotic tangents on two planes which intersect in their common cusp axis at the point.*

THEOREM 6.9. *The ray conics at a point of two reciprocal families have double contact on the asymptotic tangents at the points where the flex ray at P cuts them.*

7. Torsal curves. Let C defined by $u^2 = u^2(u^1)$ be a hypergeodesic C of (3.1) such that the direction of the curve is determined by $\lambda = du^2/du^1$. We shall determine λ so that C is a curve such that the generators corresponding to λ of the osc-cone of (3.1) at points of C form a developable surface and we shall call the curve *a torsal curve of the family* (3.1).

If C is an asymptotic line, its corresponding generators on the osc-cone are the asymptotic tangents generating a developable of which the asymptotic line is both an edge of regression and a torsal curve. Consequently an asymptotic hypergeodesic of a family (3.1) is always a torsal curve of the family and it is a plane curve if and only if it is a straight line. In the following discussion we shall exclude the asymptotic lines by assuming $\lambda \neq 0$ or ∞.

Looked upon as a curve in space the tangent at a point P of C has its direction numbers given by

$$dx^i/du^1 = \partial x^i/\partial u^a \cdot du^a/du^1 = \partial x^i/\partial u^1 + \lambda \partial x^i/\partial u^2. \tag{7.1}$$

The direction numbers of the generator of the osc-cone corresponding to C at P are given by (3.8). Let M be any point different from P on the generator. Then M may be defined by $\xi^i = x^i + \rho c^i$, where ρ is an arbitrary but fixed constant different from zero. The tangent to the curve traced by M as P moves on C has its direction numbers given by

$$\rho^{-1} d\xi^i/du^1 = \bar{P}^a \partial x^i/\partial u^a + \bar{Q} X^i, \tag{7.2}$$

where

$$\bar{P}^1 = 1/\rho + P^1\Gamma^1{}_{11} + \partial P^1/\partial u^1 + P^2\Gamma^1{}_{12} - g^{12}d_{12}R + P^1\Gamma^1{}_{12}\lambda$$
$$+ \lambda\partial P^1/\partial u^2 + P^2\Gamma^1{}_{22}\lambda - g^{11}d_{12}R\lambda,$$

$$(7.3) \quad \bar{P}^2 = \lambda/\rho + P^1\Gamma^2{}_{11} + P^2\Gamma^2{}_{12} + \partial P^2/\partial u^1 - g^{22}d_{12}R$$
$$+ P^1\Gamma^2{}_{12}\lambda + P^2\Gamma^2{}_{22}\lambda + \lambda\partial P^2/\partial u^2 - g^{12}d_{12}R\lambda,$$

$$\bar{Q} = P^2d_{12} + \partial R/\partial u^1 + P^2d_{12}\lambda + \lambda\partial R/\partial u^2,$$

which are obtained after simplification by the aid of the Gauss and the Weingarten equations.

The generator of the osc-cone will generate a developable as P moves on C if and only if the tree lines—the tangent to C, the corresponding generator of the osc-cone and the tangent to the curve traced by M—are coplanar, that is

$$(7.4) \quad e_{ijk}(\partial x^i/\partial u^\alpha \cdot du^\alpha/du^1)(P^\beta\partial x^j/\partial u^\beta + QX^j)(\bar{P}^\gamma\partial x^k/\partial u^\gamma + \bar{Q}X^k) = 0.$$

This reduces to

$$(7.5) \quad t_0 + t_1\lambda + t_2\lambda^2 + t_3\lambda^3 + t_4\lambda^4 + t_5\lambda^5 + t_6\lambda^6 = 0,$$

where a factor λ^2 is omitted and where the coefficients are defined by

$$t_0 = 3A^2 + 4A\Gamma^2{}_{11} + (\Gamma^2{}_{11})^2,$$
$$t_1 = 4AB + 2A\Gamma^2{}_{12} - 2A\Gamma^1{}_{11} + 4B\Gamma^2{}_{11} + 2\Gamma^2{}_{11}\Gamma^2{}_{12} - 2\Gamma^2{}_{11}\Gamma^1{}_{11}$$
$$+ 2(A + \Gamma^2{}_{11})(\log d_{12})_{u^1} - 2(\Gamma^2{}_{11})_{u^1} - 2A_{u^1},$$
$$t_2 = 2AC - 2A\Gamma^2{}_{22} + 2A\Gamma^1{}_{12} + 4C\Gamma^2{}_{11} - 2\Gamma^2{}_{11}\Gamma^1{}_{12} + B^2 + (\Gamma^1{}_{11})^2$$
$$+ 2B\Gamma^2{}_{12} - 2B\Gamma^1{}_{11} - 2\Gamma^2{}_{12}\Gamma^1{}_{11} + 2(B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})(\log d_{12})_{u^1}$$
$$+ 2(A + \Gamma^2{}_{11})(\log d_{12})_{u^2} - 2B_{u^1} - 4(\Gamma^2{}_{12})_{u^1} + 2(\Gamma^1{}_{11})_{u^1}$$
$$+ 4g^{22}d^2{}_{12} - 2A_{u^2} - 2(\Gamma^2{}_{11})_{u^2},$$
$$t_3 = 2(C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})(\log d_{12})_{u^1} + 2(B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})(\log d_{12})_{u^2}$$
$$+ 4D\Gamma^2{}_{11} + 2C\Gamma^2{}_{12} - 2B\Gamma^2{}_{22} - 2\Gamma^2{}_{12}\Gamma^2{}_{22} - 2B_{u^2} - 4(\Gamma^2{}_{12})_{u^2}$$
$$(7.6) \quad + 2(\Gamma^1{}_{11})_{u^2} - 2C\Gamma^1{}_{11} + 2\Gamma^1{}_{12}\Gamma^1{}_{11} - 2C_{u^1} - 2(\Gamma^2{}_{22})_{u_1}$$
$$+ 4(\Gamma^1{}_{12})_{u^1} + 2B\Gamma^1{}_{12} + 4A\Gamma^1{}_{22},$$
$$t_4 = -2BD - 2D\Gamma^1{}_{11} + 2D\Gamma^2{}_{12} + 4B\Gamma^1{}_{22} + 2\Gamma^2{}_{12}\Gamma^1{}_{22} - C^2$$
$$-(\Gamma^2{}_{22})^2 + 2C\Gamma^1{}_{12} - 2C\Gamma^2{}_{22} + 2\Gamma^2{}_{22}\Gamma^1{}_{12} + 2(D - \Gamma^1{}_{22})(\log d_{12})_{u^1}$$
$$+ 2(C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})(\log d_{12})_{u^2} - 2C_{u^2} - 2(\Gamma^2{}_{22})_{u^2}$$
$$+ 4(\Gamma^1{}_{12})_{u^2} - 4g^{11}d^2{}_{12} - 2D_{u^1} + 2(\Gamma^1{}_{22})_{u^1},$$
$$t_5 = -4DC + 2D\Gamma^1{}_{12} - 2D\Gamma^2{}_{22} + 4C\Gamma^1{}_{22} + 2\Gamma^1{}_{22}\Gamma^2{}_{22} - 2\Gamma^1{}_{22}\Gamma^1{}_{12}$$
$$+ 2(D - \Gamma^1{}_{22})(\log d_{12})_{u^2} + 2(\Gamma^1{}_{22})_{u^2} - 2D_{u^2},$$
$$t_6 = -3D^2 + 4D\Gamma^1{}_{22} - (\Gamma^1{}_{22})^2.$$

In the space in which the surface S: (2.1) is imbedded, the curve C: $u^2 = u^2(u^1)$ has equations $x^i = x^i(u^1, u^2(u^1)) = f^i(u^1)$. A necessary and sufficient condition that C be a plane curve is

$$e_{ijk} dx^i/du^1 \cdot d^2x^j/du^{12} \cdot d^3x^k/du^{13} = 0. \tag{7.7}$$

When we substitute in (7.7) the following values

$$\begin{aligned}
& dx^i/du^1 = \partial x^i/\partial u^1 + \lambda \partial x^i/\partial u^2, \\
& d^2x^j/du^{12} = p_1 \partial x^j/\partial u^1 + p_2 \partial x^j/\partial u^2 + rX^j, \\
& d^3x^k/du^{13} = \partial x^k/\partial u^1 [p_1 \Gamma^1{}_{11} + \partial p_1/\partial u^1 + p_2 \Gamma^1{}_{12} - 2d^2{}_{12} g^{12} \lambda' + p_1 \Gamma^1{}_{12} \lambda \\
& \qquad + p_2 \Gamma^1{}_{22} \lambda - 2d^2{}_{12} g^{11} \lambda^2 + \lambda \partial p_1/\partial u^2] \\
& \qquad + \partial x^k/\partial u^2 [p_1 \Gamma^2{}_{11} + p_2 \Gamma^2{}_{12} + \partial p_2/\partial u^1 - 2d^2{}_{12} g^{22} \lambda \\
& \qquad\qquad + p_1 \Gamma^2{}_{12} \lambda + p_2 \Gamma^2{}_{22} \lambda + \lambda \partial p_2/\partial u^2 - 2d^2{}_{12} g^{12} \lambda^2] \\
& \qquad + X^k [d_{12} p_2 + \partial r/\partial u^1 + p_1 d_{12} \lambda + \lambda \partial r/\partial u^2]
\end{aligned} \tag{7.8}$$

where $p_1 = \Gamma^1{}_{11} + 2\Gamma^1{}_{12}\lambda + \Gamma^1{}_{22}\lambda^2$, $p_2 = \Gamma^2{}_{11} + 2\Gamma^2{}_{12}\lambda + \Gamma^2{}_{22}\lambda^2 + \lambda'$, $r = 2d_{12}\lambda$, and simplify, we find the condition (7.7) for a hypergeodesic of (3.1) to be a plane curve is exactly the condition (7.5) for a hypergeodesic of (3.1) to be a torsal curve of (3.1). Hence

THEOREM 7.1. *Non-asymptotic hypergeodesics of a family of hypergeodesics are plane if and only if they are the torsal curves of the family. Asymptotic hypergeodesics of a family are always the torsal curves of the famliy and are plane if and only if they are straight lines.*

Equation (7.5) is in general of the sixth degree in λ and consequently has six solutions. Since λ is not zero or infinity, the vanishing of t_0 or t_6 or the vanishing of t_0, t_6 or t_0, t_1 or t_5, t_6 will reduce the degree of equation (7.5) by one or two respectively. Thus when $A = -\frac{1}{3}\Gamma^2{}_{11}$ or $D = \frac{1}{3}\Gamma^1{}_{22}$, it is in general of degree five; when $A = -\frac{1}{3}\Gamma^2{}_{11}$ and $D = \frac{1}{3}\Gamma^1{}_{22}$ or $A = -\Gamma^2{}_{11}$ or $D = \Gamma^1{}_{22}$, it is in general of degree four; when $A = -\frac{1}{3}\Gamma^2{}_{11}$ and $D = \Gamma^1{}_{22}$ or $A = -\Gamma^2{}_{11}$ and $D = \frac{1}{3}\Gamma^1{}_{22}$, it is in general of degree three; and when $A = -\Gamma^2{}_{11}$ and $D = \Gamma^1{}_{22}$, it is in general of degree two. Besides these cases in which we assume that B, C are arbitrary functions, there are other families of hypergeodesics for which B, C are also suitably chosen such that their corresponding equations (7.5) are of degree three, two, one or zero.

THEOREM 7.2. *A two-parameter family of hypergeodesics contains in general six one-parameter families of non-linear plane curves. There are*

families which contain in general five, four, three, two or one one-parameter families of such plane curves or do not contain any such plane curve. In any case, if a family contains more than the maximum number of one-parameter families of non-linear plane curve which the family is to contain in general, then all hypergeodesics of the family are plane. A family of union curves is the only family of hypergeodesics which contains in general a net of non-linear plane curves and for which B, C are arbitrary.

A comparison among the coefficients in equations (7. 5) corresponding to two reciprocal families will yield the first part of the following theorem. The second part is seen from the fact that these equations with B, C arbitrary are both of degree two when and only when $\Gamma^2{}_{11} = 0$, $\Gamma^1{}_{22} = 0$.

THEOREM 7. 3. *The numbers of one-parameter families of non-linear plane curves contained in two reciprocal families of hypergeodesics are not related. In general they are both equal to six or are unequal. There are reciprocal families such that these numbers are both equal to five, four, three, two, one or zero. When B, C are arbitrary, two reciprocal families both contain a net of non-linear plane curves if and only if the surface is a quadric.*

I am grateful to Professor P. Sperry for her interest and suggestions during the preparation of this paper.

UNIVERSITY OF CALIFORNIA, BERKELEY.

BIBLIOGRAPHY.

[1] L. P. Eisenhart, *An introduction to differential geometry with the use of the tensor calculus*, Princeton University Press (1947).

[2] G. M. Green, "Memoir on the general theory of surfaces and rectilinear congruences," *Transactions of the American Mathematical Society*, vol. 20 (1919), pp. 79-153.

[3] E. P. Lane, *Projective differential geometry of curves and surfaces*, The University of Chicago Press (1932).

[4] P. Sperry, "Properties of a certain projectively defined two-parameter family of curves on a general surface," *American Journal of Mathematics*, vol. 40 (1918), pp. 213-224.

[5] C. E. Springer, "Dual geodesics on a surface," *Bulletin of the American Mathematics Society*, vol. 48 (1942), pp. 901-906.

ON RIEMANN METRICS OF CONSTANT CURVATURE.*

By Aurel Wintner.

1. Let D be an open, simply-connected domain in the space of n real variables $x_1, \cdots, x_n$, and let (g_{ik}) be an n-rowed, real, symmetric matrix of functions $g_{ik} = g_{ik}(x_1, \cdots, x_n)$ which have a non-vanishing determinant and are of class C' (i. e., such as to possess continuous partial derivations of *first* order). Then the Christoffel symbols of second kind, $\Gamma_{ik}{}^j = \Gamma_{ik}{}^j(x_1, \cdots, x_n)$, exist and are continuous functions on D. Hence Levi-Civita's parallel transport, being determined by a linear system of ordinary differential equations with a continuous coefficient matrix, exists along every smooth path in D (here and in the sequel, a path $x_i = x_i(t)$ $(i = 1, \cdots, n)$ in D is called smooth if it has a continuously turning tangent).

The Riemannian geometry determined by (g_{ik}) on D will be called Euclidean if no vector is changed by parallel transportation along any closed smooth path. It will be proved that these metrics $ds^2 = g_{\alpha\beta} dk_\alpha dx_\beta$ are characterized by the following property:

If D is small enough, D possesses a one-to-one continuous mapping $y_i = y_i(x_1, \cdots, x_n)$ in which the n functions y_i are of class C' on D, have a non-vanishing Jacobian, and render every h_{ik} a constant, where $h_{ik} = h_{ik}(y_1, \cdots, y_n)$ is defined by placing $ds^2 = h_{\alpha\beta} dy_\alpha dy_\beta$.

The point in this theorem is that, in contrast to the classical theory, the machinery of a curvature tensor is now not made available by the geometrically artificial assumption of continuous *second* derivatives for the functions $g_{ik}(x_1, \cdots, x_n)$. In fact, since the latter are just of class C', the functions $\Gamma_{ik}{}^j(x_1, \cdots, x_n)$ can be nowhere differentiable. Correspondingly, the result seems to be new even in the case, $n = 2$, of the differential geometry of surfaces (to be embedded, rather than being given as embedded, into an (x, y, z)-space under the additional restriction that the binary form ds^2 is positive definite on D).

It will be shown at the end of the paper that the above theorem can be generalized so as to apply to those ds^2 which, when the $g_{ik}(x_1, \cdots, x_n)$ are of class C'', become the spaces of constant curvature. Since the $g_{ik}(x_1, \cdots, x_n)$ are supposed to be only of class C' (and of non vanishing determinant), the

* Received November 13, 1950.

constancy of the curvature cannot, of course, be defined in explicit terms. The substitute to be used in the C'-case is that supplied by the requirement of homogeneity, in the sense of Riemann-Helmholtz, of the metric defined by the function $g_{ik}(x_1, \cdots, x_n)$ of class C'.

In the particular case in which the x-space is Euclidean in the above-defined sense, only "the components of the affine connection," $\Gamma_{jk}{}^i = \Gamma_{kj}{}^i$, rather than the "metric tensor" g_{ik} as well, will enter into the considerations (cf. (5) below). Correspondingly, the first of the results announced above holds without the assumption of a ds^2 also. This result will be formulated and proved in Section 5.

2. Levi-Civita's system of n ordinary, linear differential equations, say $v'_i = f_{ia}(t)v_a$, is

$$v'_i = -\Gamma_{a\beta}{}^i x'_a v_\beta, \tag{1}$$

where $v_i = v_i(t)$ is the vector transported along a given smooth path $x_i = x_i(t)$ (of smooth parametrization), the prime denotes differentiation with respect to t, and all indices range from 1 to n. If the definition of a Euclidean (g_{ik}), as specified above, is compared with (1), it is readily seen that a given realization, $g_{a\beta}dx_a dx_\beta$, of a ds^2 is Euclidean on D if and only if, corresponding to every set of n numbers $v_i{}^0$ and to every point $(x_1{}^0, \cdots, x_n{}^0)$ of D, the system of n^2 linear "partial" differential equations

$$\partial v_i/\partial x_k = -\Gamma_{ka}{}^i(x_1, \cdots, x_n)v_a \tag{2}$$

(which actually is a "total" system) has exactly and/or at least one solution $v_i = v_i(x_1, \cdots, x_n)$ which is continuous, and therefore of class C', and satisfies the n initial conditions

$$v_i(x_1{}^0, \cdots, x_n{}^0) = v_i{}^0. \tag{3}$$

The standard answer to this existence question is expressed by saying that the identical vanishing of Riemann's curvature tensor (of four indices) is necessary and sufficient for the solvability of (2) under an arbitrary initial condition (3). This answer results as follows (cf., e.g., [3], pp. 22-23): A calculation shows that the curvature tensor vanishes identically if and only if

$$\partial\Gamma_{kl}{}^i/\partial x_j - \partial\Gamma_{kl}{}^i/\partial x_j = 0. \tag{4}$$

On the other hand, since the initial condition (3) is arbitrary, (4) is identical with the integrability condition, $\partial^2 v_i/\partial x_k\partial x_j = \partial^2 v_i/\partial x_j\partial x_k$, of the total system

(2). Consequently, it is sufficient to apply to (2) the general existence theorem of a total system (cf., e. g., [3], pp. 66-68) which, incidentally, is linear in the present case. But all of this depends on the assumption that the g_{ik}'s are of class C'', since otherwise the Γ's are not of class C', and therefore (4) is meaningless.

In the general case to be considered, the g_{ik}'s are of class C', hence the Γ's just continuous. What is then applicable to (2) is a recent improvement of the classical existence theorem of total systems; cf. [2]. In fact, if Theorem (II) in [2], pp. 760-761, is applied to the present particular case, (2), then the second of the two conditions (10), (11) on p. 761 in [2] reduces to $0 = 0$, while the first becomes

$$\int_J (\Gamma_{jk}{}^i dx_k + \Gamma_{jl}{}^i dx_l) = \iint_S (\Gamma_{al}{}^i \Gamma_{jk}{}^a - \Gamma_{ak}{}^i \Gamma_{jl}{}^a)\, dx_k dx_l. \tag{5}$$

Here the indices i, j and k, l are subject to the limitation $k \neq l$ but are otherwise arbitrary; J to the left denotes *any* small, rectifiable, positively oriented Jordan curve in that (x_k, x_l)-plane which results if those of the $n-2$ coordinates $x_1, \cdots, x_n$ which are distinct from x_k and x_l are arbitrarily fixed within the n-dimensional domain D, and S to the right of (5) is the two-dimensional region surrounded by J; finally, the (Greek) repeated index, a, in the double integral is, but the (Roman) repeated indices, k and l, in the line integral are not, summation indices.

Accordingly, the problem is reduced to determination of those sets of continuous Christoffel functions $\Gamma_{jk}{}^i(x_1, \cdots, x_n)$ which satisfy (5) identically (on D).

It should be noted that no reference is now allowed to the pseudo-tensor character of the $\Gamma_{jk}{}^i$. In fact, the transformation rule of the $\Gamma_{jk}{}^i$ contains the *second* derivatives of the transforming function

$$y_i = y_i(x_1, \cdots, x_n), \qquad (i = 1, \cdots, n). \tag{6}$$

3. For a single unknown function, say $y = y(x_1, \cdots, x_n)$, consider on D the system of differential equations

$$\partial^2 y/\partial x_i \partial x_k = \Gamma_{ki}{}^a(x_1, \cdots, x_n)\partial y/\partial x_a, \tag{7}$$

where i and k range from 1 to n. Clearly, (7) is equivalent to the linear "total" system which consists of

$$\partial w_k/\partial x_i = \Gamma_{ki}{}^a(x_1, \cdots, x_n) w^a \tag{8}$$

and

$$(9) \qquad \partial y/\partial x_i = w_i$$

together (with $y_1, w_1, \cdots, w_n$ as $n+1$ unknown functions of the n independent variables $x_1, \cdots, x_n$). It will be noted that (8) is the adjoint of the system (2).

Since $\Gamma_{ik}{}^a = \Gamma_{ki}{}^a$, it is readily seen that the integrability conditions, which belong to the system (8)-(9) in the same way as (5) belongs to (2) alone, are represented just by (5). Hence, if Theorem (II) of [2] is applied to (8)-(9), it follows that (5) is necessary and sufficient for the following situation:

Corresponding to any set of $n+1$ numbers a; $c_1, \cdots, c_n$, the system (7) has at least one and/or exactly one solution $y = y(x_1, \cdots, x_n)$, of class C', satisfying the initial conditions

$$(10) \qquad y(x_1{}^0, \cdots, x_n{}^0) = a, \qquad \partial y(x_1{}^0, \cdots, x_n{}^0)/\partial x_i = c_i,$$

where $(x_1{}^0, \cdots, x_n{}^0)$ is an arbitrary point of D. Actually, every such

$$(11) \qquad y(x_1, \cdots, x_n) \text{ is of class } C''.$$

In fact, if $y(x_1, \cdots, x_n)$ is of class C' and satisfies (7), then (11) follows from continuity of the coefficient functions, $\Gamma_{kj}{}^i (= \Gamma_{jk}{}^i)$, of (7).

4. Let (e_{ik}) denote the unit matrix and, under the assumption (5), let $y_k(x_1, \cdots, x_n)$, where $k = 1, \cdots, n$, denote that solution, $y = y(x_1, \cdots, x_n)$, of (7) for which the initial conditions (10) are as follows: $c_i = e_{ik}$, $a = a_k$ (say $a_k = 0$). Consider the transformation (6) determined by these n solutions of (7). The Jacobian of (6), being 1 at $(x_1{}^0, \cdots, x_n{}^0)$, cannot vanish near $(x_1{}^0, \cdots, x_n{}^0)$ (it could be concluded, by uniqueness, to be distinct from 0 throughout D, but this is immaterial here). Choose D about $(x_1{}^0, \cdots, x_n{}^0)$ so small that the inverse of the mapping (6) of D is unique. Finally, define on the y-image of D the functions h_{ik} by the assignment

$$(12) \qquad g_{\alpha\beta}(x_1, \cdots, x_n)\, dx_\alpha dx_\beta \equiv h_{\alpha\beta}(y_1, \cdots, y_n)\, dy_\alpha dy_\beta.$$

Then the matrix (h_{ik}) is constant on the y-image of D.

In view of the remarks made before (6), the last statement is quite unexpected. It goes beyond what has been announced in Section 1, where the existence of the normal form $(h_{ik}) =$ const. of (g_{ik}) was claimed only by virtue of some transformation (6) of class C'. What is now claimed is that such a transformation can be chosen of class C''. In fact, (11) shows

that the transformation (6), defined above, is of class C''. This agrees with the circumstance that, g_{ik} being a tensor, one degree of differentiability is lost by the occurrence of first derivatives in its transformation rule.

The verification of the statement, according to which $(h_{ik}) = \text{const.}$ holds by virtue of (12) and the present (6), can be omitted, precisely because the present (6) is of class C''. In fact, due to the latter circumstance, a well-known verification of $(h_{ik}) = \text{const.}$ (cf. [1], pp. 577-578) applies without any change, even though (g_{ik}) is just of class C'.

Another, more straightforward, verification of $(h_{ik}) = \text{const.}$ is contained in the more general approach to be taken in Section 5.

5. Since the integrability conditions (5), on which everything is based, contain only the Christoffel symbols, rather than the components of the metric tensor as well, it is natural to disregrad the assumption that the coefficients of (1) can be derived from a symmetric $g_{ik}(x_1, \cdots, x_n)$, of class C' and of non-vanishing determinant, by placing $\Gamma_{jk}{}^i = g^{i\alpha}\Gamma_{\alpha jk}$, $(g^{ik}) = (g_{ik})^{-1}$ and

$$2\Gamma_{ijk} = \partial g_{ij}/\partial x_k + \partial g_{ik}/\partial x_j - \partial g_{jk}/\partial x_i.$$

All that should be given is a pseudo-tensor $\Gamma_{jk}{}^i = \Gamma_{kj}{}^i$ (cf. [3], p. 12), which is assumed to be a continuous function of the position $(x_1, \cdots, x_n)$ on D. Then (1) still defines parallel transportation, and (5) is still necessary and sufficient in order that all such transportations be independent of the path.

Under the latter assumption, and without assuming the differentiability of the Γ's, it can be concluded that there exist local transformations (6), of class C'' and of non-vanishing Jacobian, which transform the functions $\Gamma_{jk}{}^i$ of $(x_1, \cdots, x_n)$ so as to make them identically 0 in $(y_1, \cdots, y_n)$. Then (2), (1) appear in the respective normal forms $\partial v_i/\partial y_k \equiv 0$, $v_i = \text{const.}$, and so the result of Section 4 for the Riemannian case follows as a corollary.

First, since (5) is assumed, Section 3 is applicable without change. Hence, it is possible to choose the n functions (6) in the same way as in Section 4. But each of these n functions satisfies (11) and (7). It follows therefore from (9) and from the transformation rule of the Γ's that the y-transform of (2) is $\partial v_i/\partial y_k \equiv 0$.

In what follows, it will again be assumed that the geometry, instead of having just an "affine connection" (Weyl), is Riemannian.

6. If λ is a real, non-vanishing constant, let (8_λ) and (7_λ) respectively denote the systems which result if the term $-\lambda g_{ik}(x_1, \cdots, x_n)y$ is added

to the right of (8) and (7). Thus (7_λ) is identical with (8_λ) by virtue of (9). As in the case $\lambda = 0$, treated above, the functions $g_{ik} = g_{ki}$ on D are supposed to be just of class C' and of non-vanishing determinant; so that the functions $\Gamma_{jk}{}^i = \Gamma_{kj}{}^i$ exist and are continuous on D.

In view of Weyl's approach to the "Helmholtz space problem" (cf. [3], pp. 24-29; also pp. 71-74), it is natural to declare the C'-metric of $g_{\alpha\beta}dx_\alpha dy_\beta$ a metric of constant curvature if, corresponding to arbitrary initial conditions (10), (3), the system consisting of (9) and (8_λ) (and belonging to a suitable $\lambda =$ const.) has a solution of class C'; in other words, if the integrability conditions of (9) and (8_λ) are satisfied.

In the C''-case, this system of integrability conditions is that representation of the curvature tensor (of four indices) which is required by formula (17) in [3], p. 26, i. e., by formula (16) in [1], p. 574. In the present C'-case, where a curvature tensor cannot be written down, the integrability conditions of (8_λ)-(9) are those integral identities, extending the above (5) from $\lambda = 0$ to any $\lambda =$ const., which result if the necessary and sufficient conditions of Theorem (II), in [2], pp. 760-761, are applied to the above system (8_λ)-(9). The resulting system of integral conditions, which differ from (5) only in additive terms containing the g_{ik}'s with the constant factor λ, will be referred to as (5_λ).

7. The following theorem can now be concluded: If $g_{\alpha\beta}(x_1, \cdots, x_n)dx_\alpha dx_\beta$ and $h_{\alpha\beta}(y_1, \cdots, y_n)dy_\alpha dy_\beta$ are two real, symmetric, n-ary forms, with coefficient functions of class C' and of non-vanishing determinants, on small x- and y-domains, respectively, and if both metrics are of constant curvature (in the sense of Section 6) and belong to the same numerical value of the scalar curvature (that is, if the number λ of Section 6 is common), then the two metrics are equivalent.

By this is meant that there exists a local transformation, (6), which is of class C', of non-vanishing Jacobian, and such as to satisfy (12). Actually, it turns out that such a transformation (6) can always be chosen to be of class C''. For the Euclidean case $(\lambda = 0)$, this was proved in Section 4. Let therefore $\lambda \neq 0$. Then the assertion can be restated as follows: If the $g_{ik}(x_1, \cdots, x_n)$ satisfy the above conditions, then, locally, there exist n functions (6), of class C'' and of non-vanishing Jacobian, which have the property that, by virtue of (6) and (12),

$$h_{ik}(y_1, \cdots, y_n) = (\lambda/y_1)^{-2}\epsilon_i e_{ik},$$

where (e_{ik}) is the unit matrix and $\epsilon_i = \pm 1$. The proof of this C''-assertion proceeds as follows:

According to Section 6, the constancy of the curvature of $g_{\alpha\beta}dx_\alpha dx_\beta$ means that (5_λ), being the integrability condition of the system (8_λ)-(9), is satisfied. But the latter system is identical with (7_λ). Hence, (7_λ) has a solution $y(x_1, \cdots, x_n)$, of class C', satisfying any given initial condition (10). Choose n linearly independent solutions, (6), for (7_λ) (as for (7) in Section 4). For the same obvious reason as in Section 3, each of these n solutions of (7_λ) will satisfy (11). Consequently, the classical C''-proof (cf. [1], pp. 579-581) of the assertion of the last formula line can be repeated without any change.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] L. Bianchi, *Vorlesungen über Differentialgeometrie*, Leipzig, 1899.

[2] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[3] H. Weyl, *Mathematische Analyse des Raumproblems*, Berlin, 1923.

ON INDUCED REPRESENTATIONS OF GROUPS.*

By George W. Mackey.[1]

Introduction. Let $\mathfrak{G}$ be a finite group, let G be a subgroup of $\mathfrak{G}$, and let $L(\xi \to L_\xi)$ be a representation of G by non-singular linear transformations in a vector space $\mathfrak{H}(L)$ which is finite-dimensional over a field $\mathfrak{F}$. Then the functions f from $\mathfrak{G}$ to $\mathfrak{H}(L)$ such that

$$f(\xi x) = L_\xi f(x) \tag{1}$$

for all $x \varepsilon \mathfrak{G}$ and all $\xi \varepsilon G$ form a vector space over $\mathfrak{F}$ which is invariant under translation from the right by members of $\mathfrak{G}$. In other words if we define U_s^L by the equation $(U_s^L(f))(x) = f(xs)$ then U^L $(s \to U_s^L)$ will be a representation of $\mathfrak{G}$ by non-singular linear transformations in the vector space $\mathfrak{H}(U^L)$ of functions f satisfying (1). The representation U^L is called the (imprimitive) representation of $\mathfrak{G}$ "induced" by L. Induced representations have been studied by many writers beginning with Frobenius, who invented them [3]. It is the purpose of part I of this article to unify a part of the work of Frobenius and his successors by showing that a number of their results are easy consequences of a single theorem (Theorem 2) about Kronecker products of induced representations. Specifically, Theorem 2 implies the Frobenius reciprocity theorem [3], Shoda's results [10] on the irreducibility and equivalence of monomial representations and Artin's theorem [1] to the effect that any character is a rational linear combination of characters of induced representations of a very special sort.

Some of our results are extensible to Hilbert space representations of locally compact topological groups. In part these extensions involve rather complicated topological and measure theoretic considerations. Certain of them however involve a minimum of such analysis and these are presented in part II. Our more involved results will appear in another paper addressed primarily to analysts.

---

* Received October 10, 1950.

[1] Most of the work on which this paper is based was done while the author was a fellow of the John Simon Guggenheim Memorial Foundation in residence successively at the University of Chicago, the University of Nancy, and the Institute for Advanced Study.

I. Finite Groups.

1. Definitions and preliminary lemmas. We shall adhere to the notation introduced in the introduction and in particular shall always use $\mathfrak{H}(U)$ to denote the vector space of a representation U. If $\mathfrak{H}$ is any vector space over $\mathfrak{F}$ we shall denote by $\bar{\mathfrak{H}}$ the conjugate space; that is the space of all linear functions from $\mathfrak{H}$ to $\mathfrak{F}$. If T is a linear transformation of $\mathfrak{H}$ into itself, T^* will denote the linear transformation of $\bar{\mathfrak{H}}$ into itself such that $T^*(l)(v) = l(T(v))$ for all l in $\bar{\mathfrak{H}}$ and all v in $\mathfrak{H}$.

If $U(x \to U_x)$ is any representation of $\mathfrak{G}$ we shall denote by $\bar{U}$ and call the *adjoint* of U the representation $x \to U_{x^{-1}}^*$. Clearly $\mathfrak{H}(\bar{U}) = \overline{\mathfrak{H}(U)}$.

Let $U(x \to U_x)$ and $V(y \to V_y)$ be representations of the groups $\mathfrak{G}_1$ and $\mathfrak{G}_2$. Let $\mathfrak{H}_1$ be the set of all linear transformations from $\overline{\mathfrak{H}(V)}$ to $\mathfrak{H}(U)$. For each $x, y \,\varepsilon\, \mathfrak{G}_1 \times \mathfrak{G}_2$, $T \to U_x T V_y^*$ is a linear transformation of $\mathfrak{H}_1$ into itself. If we denote this by $U_x \times V_y$ then $x, y \to U_x \times V_y$ is a representation of $\mathfrak{G}_1 \times \mathfrak{G}_2$ in $\mathfrak{H}_1$. We shall call this representation of $\mathfrak{G}_1 \times \mathfrak{G}_2$ the *outer Kronecker product* $U \times V$ of U and V. Now let U and V be representations of the same group $\mathfrak{G}$. $\mathfrak{G} \times \mathfrak{G}$ then contains a subgroup $\tilde{\mathfrak{G}}$ isomorphic to $\mathfrak{G}$; namely the group of all x, y with $x = y$. $U \times V$ restricted to $\tilde{\mathfrak{G}}$ then defines a representation of $\mathfrak{G}$. We call this the *Kronecker product* $U \otimes V$ of U and V.

Let L be a representation of the subgroup G of $\mathfrak{G}$ and consider $\mathfrak{H}(U^L)$. Let l' be any member of $\overline{\mathfrak{H}(U^L)}$. Then as is easy to see there exists a function l'' $(x \to l_x'')$ from $\mathfrak{G}$ to $\overline{\mathfrak{H}(L)}$ such that $l'(f) = \Sigma_{x \varepsilon \mathfrak{G}} l_x''(f(x))$ for all $f \,\varepsilon\, \mathfrak{H}(U^L)$. But $\Sigma_{x \varepsilon \mathfrak{G}} l_x''(f(x)) = \Sigma_{\mathfrak{G}/G} \Sigma_{\xi \varepsilon G} l_{\xi x}''(f(\xi x))$, where the second sum is over the right-G-cosets. Now

$$\Sigma_{\xi \varepsilon G} l_{\xi x}''(f(\xi x)) = \Sigma_{\xi \varepsilon G} L_\xi^*(l_{\xi x}'')(f(x)).$$

Thus if we set $l_x = \Sigma_{\xi \varepsilon G} L_\xi^*(l_{\xi x}'')$ we have $l_{\xi x} = L_{\xi^{-1}}^*(l_x)$ for all $\xi \,\varepsilon\, G$ and all $x \,\varepsilon\, \mathfrak{G}$ and

$$l(f) = \Sigma_{\mathfrak{G}/G} l_x(f(x)). \tag{2}$$

The reader will now have no difficulty in verifying the truth of

LEMMA 1. *The mapping of $\mathfrak{H}(U^{\bar{L}})$ into $\overline{\mathfrak{H}(U^L)}$ defined by equation (2) is an isomorphism onto and sets up an equivalence between the representations $U^{\bar{L}}$ and $\overline{U^L}$.*

Next let G_1 and G_2 be subgroups $\mathfrak{G}_1$ and $\mathfrak{G}_2$ respectively and let L and M be representations of G_1 and G_2. Let A be any member of $\mathfrak{H}(U^{L \times M})$. Then A is a function from $\mathfrak{G}_1 \times \mathfrak{G}_2$ to $\mathfrak{H}(L \times M)$ such that

$A_{\xi x,\eta y} = (L_\xi \times M_\eta)(A_{x,y})$ for all $x, y \in \mathfrak{G}_1 \times \mathfrak{G}_2$ and all $\xi, \eta \in G_1 \times G_2$. In other words $A_{x,y}$ is for each x, y a linear transformation from $\overline{\mathcal{H}(M)}$ to $\mathcal{H}(L)$ and $A_{\xi x,\eta y} = L_\xi A_{x,y} M_\eta^*$ for all $\xi, \eta \in G_1 \times G_2$ and all $x, y \in \mathfrak{G}_1 \times \mathfrak{G}_2$. Thus if l $(y \to l_y)$ is any member of $\mathcal{H}(U^{\bar{M}})$ we see that $A_{x,y}(l_y)$ as a function of y depends only upon the right G_2-coset to which y belongs and that we may define a member f of $\mathcal{H}(U^L)$ as follows:

$$f(x) = \sum^{y}_{\mathfrak{G}_2/G_2} A_{x,y}(l_y). \tag{3}$$

By Lemma 1 there is a natural map of $\mathcal{H}(U^{\bar{M}})$ on $\overline{\mathcal{H}(U^M)}$. Thus (3) may be regarded as defining a linear map of $\overline{\mathcal{H}(U^M)}$ on $\mathcal{H}(U^L)$; that is, a member of $\mathcal{H}(U^L \times U^M)$. Let us denote this member of $\mathcal{H}(U^L \times U^M)$ by $A^\sim$. The reader can now readily satisfy himself of the truth of

LEMMA 2. *The mapping $A \to A^\sim$ of $\mathcal{H}(U^{M\times L})$ into $\mathcal{H}(U^L \times U^M)$ is an isomorphism onto and sets up an equivalence between the representations $U^L \times U^M$ and $U^{L\times M}$.*

2. The main theorems. Let now G_1 and G_2 be two subgroups of the group $\mathfrak{G}$ and let L be a representation of G_1. Let us denote by ${}_{G_2}U^L$ the restriction to G_2 of the representation U^L of $\mathfrak{G}$. We observe at once that ${}_{G_2}U^L$ admits a direct sum decomposition into as many parts as there are double cosets G_1xG_2. Indeed let $x_1, x_2, \cdots, x_n$ be a choice of elements, one from each double coset, so that $\mathfrak{G} = \bigcup_{i=1}^{n} G_1x_iG_2$. For each $i = 1, 2, \cdots, n$ and each $f \in \mathcal{H}(U^L)$ let $f_i(x) = f(x)$ if $x \in G_1x_iG_2$ and let $f_i(x) = 0$ if $x \notin G_1x_iG_2$. Then f_i is also in $\mathcal{H}(U^L)$ and the set of all f_i for $f \in \mathcal{H}(U^L)$ is precisely the set $\mathcal{H}_i$ of all $f \in \mathcal{H}(U^L)$ which vanish outside of $G_1x_iG_2$. Finally it is clear that for each $s \in G_2$, U_s^L carries each $\mathcal{H}_i$ into itself. Thus $\mathcal{H}(U^L) = \mathcal{H}_1 \oplus (\mathcal{H}_2 \oplus \cdots \oplus \mathcal{H}_n$ defines a direct sum decomposition of ${}_{G_2}U^L$. Let us examine the representation associated with $\mathcal{H}_i$. If f is any member of $\mathcal{H}_i$ we may define a function $\tilde{f}$ from G_2 to $\mathcal{H}(L)$ as follows: $\tilde{f}(t) = f(x_it)$. Since $f(\xi x_it) = L_\xi f(x_it) = L_\xi \tilde{f}(t)$ for all $\xi \in G_1$ and $t \in G_2$, it is clear that $\tilde{f}$ determines f completely and that $f \to \tilde{f}$ is a one-to-one map. Let $x_i^{-1}\xi x_i = \eta$ be any element of $G_2 \cap x_i^{-1}G_1x_i$. Then $\tilde{f}(\eta t) = f(\xi x_it) = L_\xi f(x_it) = L_{x_i\eta x_i^{-1}}(\tilde{f}(t))$. Conversely, if g is any function from G_2 to $\mathcal{H}(L)$ such that $g(\eta t) = L_{x_i\eta x_i^{-1}}(g(t))$ for all $\eta \in G_2 \cap x_i^{-1}G_1x_i$ and all $t \in G_2$, then it may be verified at once that $f(\xi x_it) = L_\xi(g(t))$ defines f unambiguously as a member of $\mathcal{H}_i$ such that $g = \tilde{f}$. In other words $f \to \tilde{f}$ sets up an equivalence between the component of ${}_{G_2}U^L$ in $\mathcal{H}_i$ and the representation of G_2 induced by the representation $\eta \to L_{x_i\eta x_i^{-1}}$ of the subgroup $G_2 \cap x_i^{-1}G_1x_i$. Thus we have proved

THEOREM 1. *Let L, G_1, G_2 and $\mathfrak{G}$ be described above. For each $x \varepsilon \mathfrak{G}$ consider the subgroup $G_2 \cap x^{-1}G_1x$ of G_2 and let V^x denote the representation of G_2 induced by the representation $\eta \to L_{x\eta x^{-1}}$ of this subgroup. Then V^x is determined to within equivalence by the double coset $G_1xG_2 = D(x)$ to which x belongs and we may write $V^D = V^x$ where $D = D(x)$. U^L restricted to G_2 is the direct sum of the V^D over the double $G_1 : G_2$ cosets D.*

As an almost immediate consequence of Theorem 1 and Lemma 2 we have

THEOREM 2.[2] *Let G_1 and G_2 be subgroups of $\mathfrak{G}$ and let L and M be representations of G_1 and G_2 respectively. For each $x, y \varepsilon \mathfrak{G} \times \mathfrak{G}$ consider the representations $s \to L_{xsx^{-1}}$ and $s \to M_{ysy^{-1}}$ of the subgroup $(x^{-1}G_1x) \cap (y^{-1}G_2y)$ of $\mathfrak{G}$. Let us denote their Kronecker product by $N^{x,y}$ and form the induced representation $U^{N^{x,y}}$ of $\mathfrak{G}$. Then $U^{N^{x,y}}$ is determined to within equivalence by the double coset $G_1xy^{-1}G_2$ to which xy^{-1} belongs and the direct sum of the $U^{N^{x,y}}$ over the double cosets is equivalent to the Kronecker product $U^L \otimes U^M$ of U^L and U^M.*

Proof. $U^L \otimes U^M$ is the representation of $\mathfrak{G}$ obtained from the representation $U^L \times U^M$ of $\mathfrak{G} \times \mathfrak{G}$ by restriction to the isomorphic replica $\tilde{\mathfrak{G}}$ of $\mathfrak{G}$ consisting of all $x, y \varepsilon \mathfrak{G} \times \mathfrak{G}$ with $x = y$. Moreover by Lemma 2, $U^L \times U^M$ is equivalent to $U^{L \times M}$, where $L \times M$ is of course a representation of $G_1 \times G_2$. By Theorem 1, $U^{L \times M}$ restricted to $\tilde{\mathfrak{G}}$ is a direct sum over the double cosets $(G_1 \times G_2)(x, y)\tilde{\mathfrak{G}}$ and the summand associated with the double coset containing x, y is the representation of $\tilde{\mathfrak{G}}$ induced by the representation $s, s \to (L \times M)_{(x,y)(s,s)(x,y)^{-1}}$ of $\tilde{\mathfrak{G}} \cap (x, y)^{-1}(G_1 \times G_2)(x, y)$. But it is easily verified that $\tilde{\mathfrak{G}} \cap (x, y)^{-1}(G_1 \times G_2)(x, y)$ transferred to $\mathfrak{G}$ by the natural isomorphism is the subgroup $x^{-1}G_1x \cap y^{-1}G_2y$ and the the representation $s, s \to (L \times M)_{(x,y)(s,s)(x,y)^{-1}}$ becomes $s \to L_{xsx^{-1}} \otimes M_{ysy^{-1}}$. Finally a simple calculation shows that $x_1, y_1 \varepsilon (G_1 \times G_2)(x, y)\tilde{\mathfrak{G}}$ if and only if $x_1y_1^{-1} \varepsilon G_1xy^{-1}G_2$.

Given two representations U and V of $\mathfrak{G}$ let us define the intertwining number $\mathcal{I}(U, V)$ of U and V to be the dimension of the vector space of all linear transformations T from $\mathcal{H}(V)$ to $\mathcal{H}(U)$ such that $U_xT = TV_x$ for all $x \varepsilon \mathfrak{G}$. As is evident and well known, $\mathcal{I}(U, V)$ is equal to the number of times that the Kronecker product $U \otimes V$ contains the identity; that is to

[2] We are indebted to the referee for pointing out that the special case of Theorem 2 in which $G_2 = \mathfrak{G}$ and L is the identity representation of G_1 has been obtained by M. Osima [9]. The referee reports further that Prof. Osima later obtained more general results—one he believes quite close to ours—but that so far as he knows these results have not been published.

the dimension of the set of all $w \varepsilon \mathfrak{H}(U \otimes V)$ such that $(U \otimes V)_x(w) = w$ for all $x \varepsilon \mathfrak{G}$. Moreover for any induced representation U^L the number of times that U^L contains the identity is equal to the number of times that L contains the identity. Indeed, if $f \varepsilon \mathfrak{H}(U^L)$ then $U^L(f) = f$ if and only if f is a constant function on $\mathfrak{G}$, and if $f(x) = w$ for all $x \varepsilon \mathfrak{G}$ then $f \varepsilon \mathfrak{H}(U^L)$ if and only if $L_\xi(w) = w$ for all $\xi \varepsilon G_1$. From these remarks and Theorem 2 we deduce at once

THEOREM 3. *Let $\mathfrak{G}$, G_1, G_2, L and M be as in Theorem 2. For each $x, y \varepsilon \mathfrak{G}$ consider the representations $s \to L_{xsx^{-1}}$ and $s \to M_{ysy^{-1}}$ of $x^{-1}G_1x \cap y^{-1}G_2y$, and let $\mathfrak{J}(L, M, x, y)$ denote their intertwining number. Then $\mathfrak{J}(L, M, x, y)$ depends only upon the double coset $D = D(x, y) = G_1xy^{-1}G_2$ to which xy^{-1} belongs, so that we may write $\mathfrak{J}(L, M, D)$. Finally $\mathfrak{J}(U^L, U^M) = \Sigma_{D \varepsilon \mathfrak{D}} \mathfrak{J}(L, M, D)$, where $\mathfrak{D}$ is the class of $G_1 : G_2$ double cosets.*

3. Applications. Our applications of Theorem 3 depend upon the following well-known property of intertwining numbers. Let U and V be representations of $\mathfrak{G}$ such that U is irreducible and let $\mathfrak{F}$ be algebraically closed and such that its characteristic does not divide the order of $\mathfrak{G}$. Then V is a direct sum of irreducible parts and $\mathfrak{J}(U, V)$ is equal to the number of parts which are equivalent to U. We deduce at once

THEOREM 4. (Frobenius) *Let G_1 be a subgroup of $\mathfrak{G}$ and let L and M be irreducible representations of G_1 and $\mathfrak{G}$ respectively. Let $\mathfrak{F}$ be algebraically closed and of characteristic prime to the order of $\mathfrak{G}$. Then the number of times that U^L contains M as a direct summand is equal to the number of times that M restricted to G_1 contains L as a direct summand.*

Proof. In Theorem 3 take $G_2 = \mathfrak{G}$ and observe that there is only one double coset.

As a corollary of Theorems 3 and 4 we have

THEOREM 5. (Frobenius) *Let $\mathfrak{F}$ be as in Theorem 4 and let G_1 and G_2 be subgroups of $\mathfrak{G}$. For each $i = 1, 2$ and each irreducible representation V^γ $(\gamma = 1, 2, \cdots, \sigma)$ of $\mathfrak{G}$ let $r_\gamma{}^i$ be the number of times that V^γ restricted to G_i contains the identity representation. Then $\Sigma_{\gamma=1}^{\sigma} r_\gamma{}^1 r_\gamma{}^2$ is equal to the number of $G_1 : G_2$ double cosets.*

Proof. By Theorem 4, $\Sigma_{\gamma=1}^{\sigma} r_\gamma{}^1 r_\gamma{}^2 = \mathfrak{J}(U^L, U^M)$, where L (resp. M)

is the identity representation of G_1 (resp. G_2). By Theorem 3, $\mathfrak{I}(U^L, U^M)$ is in this case the number of $G_1 : G_2$ double cosets.

THEOREM 6. *Let $\mathfrak{G}$, G_1, and $\mathfrak{F}$ be as in Theorem 4. Then U^L is irreducible if and only if for each x not in G_1 the representations of $x^{-1}G_1x \cap G_1$ defined by $s \rightarrow L_{xsx^{-1}}$ and $s \rightarrow L_s$ have no common irreducible components.*

Proof. Observe that a representation is irreducible if and only if its intertwining number with itself is 1 and apply Theorem 3 with $G_2 = G_1$ and $L = M$.

Theorem 6 for the special case in which L is one-dimensional is due to Shoda [10]. In this case its statement is slightly simpler for the two representations $s \rightarrow L_{xsx^{-1}}$ and $s \rightarrow L_s$ are one-dimensional and hence fail to have common irreducible components if and only if they are distinct. Our next corollary of Theorem 3 is also due to Shoda [10] in the special case in which L and M are one-dimensional.

THEOREM 7. *Let $\mathfrak{G}$, G_1, G_2, L and M be as in Theorem 2 and let $\mathfrak{F}$ be as in Theorem 4. Let U^L and U^M be irreducible. Then U^L and U^M are equivalent if and only if there exists $x \varepsilon \mathfrak{G}$ such that the representations $s \rightarrow L_{xsx^{-1}}$ and $s \rightarrow M_s$ of the subgroup $x^{-1}G_1x \cap G_2$ have an irreducible component in common.*

For our next application we shall need the stronger Theorem 2. We begin with a definition and a lemma. Let G_1 be a cyclic subgroup of $\mathfrak{G}$ and let L be a one-dimensional representation of G_1. Then we shall call U^L an Artin representation of $\mathfrak{G}$ and its character $\chi(U^L, x) = \text{Trace}\,(U_x{}^L)$ an Artin character.

LEMMA. *If $\mathfrak{F}$ is as in Theorem 4 and x_1 and x_2 are in different conjugate classes in $\mathfrak{G}$ then there exists an Artin character χ such that $\chi(x_1) \neq \chi(x_2)$.*

Proof. Let G_1 be the cyclic group generated by x_1, let its order be n and let ω be a primitive n-th root of unity. Then the general one-dimensional representation of G_1 is obtained by choosing $k = 0, 1, \cdots, n-1$ and setting $L_k(x_1{}^r) = \omega^{kr}$ for $r = 1, 2, \cdots, n$. Let the classes containing x_1 and x_2 be denoted by C_1 and C_2 and let C_1, C_2 and $\mathfrak{G}$ contain h_1, h_2 and h elements respectively. From the well-known and easily verified formula $\chi(U^L, x) = (1/n)\Sigma_{s \varepsilon G}\phi(s^{-1}xs)$, where $\phi(\xi) = \chi(L, \xi)$ or zero according as ξ is or is not in G_1 and where L is a representation of G_1, we compute at once that

$$\chi(U^{L_k}, x_1) = (1/n)(h/h_1)\Sigma_{x_1^r \varepsilon C_1}\omega^{kr} = \Sigma_{x_1^r \varepsilon C_1}(h/h_1 n)\omega^{kr}$$

and

$$\chi(U^{L_k}, x_2) = (1/n)(h/h_2)\Sigma_{x_1^r \varepsilon C_2}\omega^{kr} = \Sigma_{x_1^r \varepsilon C_2}(h/h_2 n)\omega^{kr}.$$

But the mappings $\theta \to \psi$, where $\psi(k) = \Sigma_{r=0}^{n-1}\theta(r)\omega^{kr}$, and $\psi \to \theta$, where $\theta(r) = \Sigma_{k=0}^{n-1}\omega^{-kr}\psi(k)$, set up a one-to-one correspondence between $\mathfrak{F}$-valued functions θ and $\mathfrak{F}$-valued functions ψ. Thus $\chi(U^{L_k}, x_1) \neq \chi(U^{L_k}, x_2)$ for at least one k. This establishes the truth of the lemma.

As a consequence of this lemma and Theorem 2 we may now prove

THEOREM 8. (Artin) *If $\mathfrak{F}$ is as in Theorem 4 then any character χ on $\mathfrak{G}$ is a finite linear combination with rational coefficients of Artin characters.*

Proof. Let $\mathfrak{A}$ be the set of all class functions on $\mathfrak{G}$ which are finite linear combinations with coefficients in $\mathfrak{F}$ of Artin characters. It is an immediate corollary of Theorem 2 that the Kronecker product of two Artin representations is a direct sum of Artin representations and hence that the product of two Artin characters is a linear combination with positive integral coefficients of Artin characters. Thus $\mathfrak{A}$ under multiplication is an algebra over $\mathfrak{F}$. Moreover by the immediately preceding lemma it contains functions distinguishing between any two classes. It follows at once that $\mathfrak{A}$ contains all class functions with values in $\mathfrak{F}$. Thus if there are c classes in $\mathfrak{G}$ there must exist c Artin characters $\chi_1, \chi_2, \cdots, \chi_c$ which form a basis for $\mathfrak{A}$ over $\mathfrak{F}$. Let $\chi_1^0, \chi_2^0, \cdots, \chi_c^0$ be the irreducible characters of $\mathfrak{G}$. Then we have $\chi_i = \Sigma_{j=1}^{c} n_{ij}\chi_j^0$, where $\| n_{ij} \|$ is a matrix of integers. Since the χ_i form a basis, $\| n_{ij} \|$ has an inverse $\| r_{ij} \|$. Since the n_{ij} are integers, the r_{ij} are rational and the truth of the theorem is evident.

It is to be remarked that Artin proved the somewhat stronger result that the only prime factors occurring in the denominators of the r_{ij} are divisors of h and that Brauer [2] has since shown that by admitting certain non-cyclic subgroups one can dispense altogether with non-integral rational numbers. Whether the methods of this paper can be used to obtain new proofs of these results is not known.

II. Infinite Groups.

1. Preliminaries. From now on the group $\mathfrak{G}$ will be a separable locally compact topological group. By a unitary representation U $(x \to U_x)$ of $\mathfrak{G}$ we shall mean a homomorphism of $\mathfrak{G}$ into the group of all unitary trans-

formations of some Hilbert space $\mathfrak{H}(U)$ into itself. We shall consider only unitary representations which are continuous in the sense that for each $v \varepsilon \mathfrak{H}(U)$ the mapping $x \to U_x(v)$ from $\mathfrak{G}$ to $\mathfrak{H}(U)$ is continuous and we shall use the words "unitary representation" to mean continuous unitary representation. We remark that in order to ensure the continuity of a U it is enough to ensure that for each v and w in $\mathfrak{H}(U)$ the function $U_x(v) \cdot w$ of x is a Borel function. Here the dot indicates the scalar product in the Hilbert space. Two unitary representations U and V will be said to be unitary equivalent if there exists a unitary transformation A from $\mathfrak{H}(U)$ to $\mathfrak{H}(V)$ such that $AU_xA^{-1} = V_x$ for all x in $\mathfrak{G}$.

The definition of U^L where L is a unitary representation of the closed subgroup G of $\mathfrak{G}$ is rather complicated in the general case. See [5]. Here we shall have occasion to deal only with subgroups G which are open as well as closed so that the space $\mathfrak{G}/G$ of right-G-cosets is denumerable and discrete. In this case we may define U^L as follows. Consider first the vector space of all functions f from $\mathfrak{G}$ to $\mathfrak{H}(L)$ such that $f(\xi x) = L_\xi f(x)$ for all $\xi \varepsilon G$ and $x \varepsilon \mathfrak{G}$. For each such f we have $f(\xi x) \cdot f(\xi x) = L_\xi f(x) \cdot L_\xi f(x)$ $= f(x) \cdot f(x)$ since L_ξ is unitary. Thus $f(x) \cdot f(x)$ depends only upon the right-G-coset to which x belongs. Let us set $\| f \|^2 = \Sigma_{G/G} f(x) \cdot f(x)$, where the sum is over the right-G-cosets. We define $\mathfrak{H}(U^L)$ to be the set of all f in the space under consideration for which $\| f \|^2 < \infty$. We leave to the reader the routine task of verifying that $\mathfrak{H}(U^L)$ is indeed a Hilbert space under the norm just defined, that $f \cdot g = \Sigma_{G/G} f(x) \cdot g(x)$, and that if $U_s^L(f)(x) = f(xs)$ then $s \to U_s^L$ is a unitary representation of $\mathfrak{G}$.

Consider now the problem of carrying over the theory of part I. If we try to do this more or less directly we are confronted with a dilemma. Suppose we define the space of the Kronecker produce of U and V as the set of all bounded linear operators from $\overline{\mathfrak{H}(V)}$ to $\mathfrak{H}(U)$. Then not only do we not get a Hilbert space but we find that the obvious generalizations of Lemma 2 and Theorem 2 are false. Suppose on the other hand that we secure a Hilbert space for $\mathfrak{H}(U \otimes V)$ by restricting attention to those operators form $\overline{\mathfrak{H}(V)}$ to $\mathfrak{H}(U)$ which are in the Schmidt class. Then we may prove analogues of Theorems 1 and 2 and even of Theorem 3 if we define intertwining numbers using only operators in the Schmidt class. Unfortunately, however, an analogue of Theorem 3 involving such "strong" intertwining numbers does not have implications about irreducibility and unitary equivalence. Thus when we pass to the infinite case we lose some of the unity of part I and must consider separately Kronecker products and restrictions to subgroups on the one hand and irreducibility, unitary equi-

valence and ordinary intertwining numbers on the other. Here we shall treat only the second topic and that in the special case in which the subgroups under consideration are both open and closed. It turns out to be possible, in this case at least, to prove an analogue of Theorem 3 without using Kronecker products. Our results will be found summarized in [6]. Our results on the first topic will be treated in detail elsewhere. They are summarized in [7].

2. The main theorem. As suggested above we define an intertwining operator T for the representations U and V of $\mathfrak{G}$ as a bounded linear operator from $\mathfrak{H}(V)$ to $\mathfrak{H}(U)$ such that $U_xT = TV_x$ for all x in $\mathfrak{G}$. The intertwining number $\mathfrak{J}(U, V)$ is then defined as the dimension $(= 0, 1, 2, \cdots, \infty)$ of the vector space of all such intertwining operators. We wish to compute $\mathfrak{J}(U^L, U^M)$ where L and M are unitary representations of the open closed subgroups G_1 and G_2 of $\mathfrak{G}$.

Suppose for the moment that $\mathfrak{G}$ is finite and that the representations considered are finite-dimensional, so that the methods of part I may be applied. The fact that the intertwining operators of U^L and U^M are defined by members of the space of $U^L \otimes \overline{U^M}$ and that the latter is an isomorphic image of the space of $U^{L\times\bar{M}}$ tells us that every intertwining operator is defined by a function A from $\mathfrak{G} \times \mathfrak{G}$ to the space of all linear transformations from $\mathfrak{H}(M)$ to $\mathfrak{H}(L)$ such that $A_{\xi x,\eta y} = L_\xi A_{x,y} M_\eta{}^*$ for all $\xi,\eta \in G_1 \times G_2$. Conversely, every such function A defines a linear transformation from $\mathfrak{H}(U^M)$ to $\mathfrak{H}(U^L)$ as follows. Consider $A_{x,y}(f(y))$, where $f \in \mathfrak{H}(U^M)$. Then for each $\eta \in G_2$, $A_{x,\eta y}(f(\eta y)) = A_{x,y}M_\eta{}^* M_\eta(f(y)) = A_{x,y}(f(y))$. Thus for fixed x, $A_{x,y}(f(y))$ depends only upon the right-G_2-coset to which y belongs. We may sum over these right-cosets and indicate this by $\overset{y}{\Sigma}_{G/G_2} A_{x,y}(f(y))$. The resulting function of x is in $\mathfrak{H}(U^L)$. The transformation $A^\sim$ from $\mathfrak{H}(U^M)$ to $\mathfrak{H}(U^L)$ thus defined is an intertwining operator if and only if $A_{xs,ys} = A_{x,y}$ for all $s \in \mathfrak{G}$. In short $\mathfrak{J}(U^L, U^M)$ is the dimension of the set of all functions A such that $A_{\xi xs,\eta ys} = L_\xi A_{x,y} M_\eta{}^*$ for all $\xi \in G_1$, $\eta \in G_2$, $s \in \mathfrak{G}$. Theorem 3 follows fairly readily from this fact.

Let us return now to the infinite case. Using the openness of G_1 and G_2 it is easy to show that every intertwining operator is defined by a function A on $\mathfrak{G} \times \mathfrak{G}$ of the sort just described. Unfortunately not every such function defines an intertwining operator and it has not proved possible to characterize those which do in a useful manner. There does turn out however to exist a conveniently characterizable subclass of functions A which can

be proved to have the same dimension as the class of those which do define intertwining operators. Our analogue of Theorem 3 follows from a determination of the dimension of this class. We proceed now to the details of these matters.

LEMMA A. *Let T be any intertwining operator for the unitary representations U^L and U^M of $\mathfrak{G}$; L and M being unitary representations of the open and closed subgroups G_1 and G_2 of $\mathfrak{G}$. Then there exists a function A $(x, y \to A_{x,y})$ from $\mathfrak{G} \times \mathfrak{G}$ to the class of bounded linear operators from $\mathfrak{H}(M)$ to $\mathfrak{H}(L)$ such that*

(a) $A_{\xi xs, \eta ys} = L_\xi A_{x,y} M_\eta^*$ *for all* $\xi \varepsilon G_1$, $\eta \varepsilon G_2$, $s \varepsilon \mathfrak{G}$.

(b) *there exists a constant* $K > 0$ *such that*

$$\overset{x}{\Sigma}_{G/G_1} \| A_{x,y}(v) \|^2 / \| v \|^2 \leq K \text{ for all } y \varepsilon \mathfrak{G} \text{ and all } v \varepsilon \mathfrak{H}(M),$$

(c) *there exists a constant* $K' > 0$ *such that*

$$\overset{y}{\Sigma}_{G/G_2} \| A_{x,y}^*(v) \|^2 / \| v \|^2 \leq K' \text{ for all } x \varepsilon \mathfrak{G} \text{ and all } v \varepsilon \mathfrak{H}(L),$$

(d) *for each* $f \varepsilon \mathfrak{H}(U^M)$, $(T(f))(x) = \overset{y}{\Sigma}_{G/G_2} A_{x,y}(f(y))$,

(e) *for each* $f \varepsilon \mathfrak{H}(U^L)$, $(T^*(f))(x) = \overset{x}{\Sigma}_{G/G_1} A_{x,y}^*(f(y))$,

the summation in (b), (c), (d), *and* (e) *being over the right-cosets indicated.*

Proof. Given an element y of $\mathfrak{G}$ and an element v of $\mathfrak{H}(M)$, let $f_{y,v}(\eta y) = M_\eta(v)$ for all $\eta \varepsilon G_2$ and let $f_{y,v}(x) = 0$ for all x not in the right-coset $G_2 y$. Then $f_{y,v}$ is clearly a member of $\mathfrak{H}(U^M)$. Moreover any member g of $\mathfrak{H}(U^M)$ which is zero outside of $G_2 y$ is of the form $f_{y,v}$ with $v = g(y)$. Finally every member of $\mathfrak{H}(U^M)$ is a unique sum over the right-G_2-cosets of members vanishing outside of these cosets. Thus the intertwining operator T is completely determined by its values $T(f_{y,v})$ at the $f_{y,v}$. Consider now $(T(f_{y,v}))(x)$. It is a member of $\mathfrak{H}(L)$ which for fixed x and y depends linearly on v. Let us denote this linear transformation from $\mathfrak{H}(M)$ to $\mathfrak{H}(L)$ by $A_{x,y}$, so that $A_{x,y}(v) = (T(f_{y,v}))(x)$. Now $\| f_{y,v} \| = \| v \|$, so $\| T(f_{y,v}) \| \leq \| T \| \| v \|$. Thus $\| A_{x,y}(v) \| \leq \| T \| \| v \|$ for all x and y, and $A_{x,y}$ is seen to be a bounded linear transformation. We shall leave to the reader the straightforward computations necessary to verify equations (a), (d) and (e). Inequality (b) is an immediate consequence of the fact that T is bounded and that $T(f_{y,v})$ is in $\mathfrak{H}(U^L)$. Indeed, $\| T(f_{y,v}) \|^2$

$= \Sigma_{G/G_1} \| A_{x,y}(v) \|^2 \leq \| T \|^2 \| f_{y,v} \|^2 = \| T \|^2 \| v \|^2$ and we may take $K = \| T \|^2$. Inequality (c) follows on applying the same argument to T^*.

LEMMA B. *$\mathfrak{J}(U^L, U^M)$ is equal to the dimension d of the vector space of all functions A having the properties* (a), (b) *and* (c) *listed in Lemma* A.

Proof. It follows at once from Lemma A that $d \geq \mathfrak{J}(U^L, U^M)$. Thus if $\mathfrak{J}(U^L, U^M) = \infty$ the equality holds trivially. It will suffice then to prove that whenever A satisfies (a), (b) and (c) and $\mathfrak{J}(U^L, U^M) < \infty$ then A defines an intertwining operator via equation (d). Let A have these properties. Let $\mathfrak{M}$ be the subspace of $\mathfrak{H}(U^M)$ consisting of all $f \varepsilon \mathfrak{H}(U^M)$ which vanish outside of a finite number of right-G_2-cosets. $\mathfrak{M}$ is dense in $\mathfrak{H}(U^M)$. Since all but a finite number of the terms vanish there is no difficulty in forming the sum $\overset{y}{\Sigma}_{G/G_2} A_{x,y}(f(y))$ for each $f \varepsilon \mathfrak{M}$. This sum as a function of x is then an element of $\mathfrak{H}(U^L)$. We obtain in this way a linear operator T_0 defined in $\mathfrak{M} \subseteq \mathfrak{H}(U^M)$ and having values in $\mathfrak{H}(U^L)$. Similarly $g \rightarrow \overset{x}{\Sigma}_{G/G_1} A_{x,y}{}^*(g(x))$ defines a linear operator T_1 whose arguments form a dense subspace $\mathfrak{L}$ of $\mathfrak{H}(U^L)$ and whose values are in $\mathfrak{H}(U^M)$. We verify at once that T_0 and T_1 are adjoint to one another; that is, $T_0(f) \cdot g = f \cdot T_1(g)$ for all $g \varepsilon \mathfrak{L}$ and $f \varepsilon \mathfrak{M}$. Hence $T_0{}^{**}$ exists and is a closed linear operator extending T_0. Moreover $T_0{}^*$ is a closed linear operator extending T_1. Now it follows at once from the definition of T_0 that $U_x{}^L T_0 = T_0 U_x{}^M$ for all $x \varepsilon \mathfrak{G}$. Hence

$$(*) \qquad T_0{}^* \overset{L}{U}_{x^{-1}} = \overset{M}{U}_{x^{-1}} T_0{}^*$$

$$(**) \qquad U_x{}^L T_0{}^{**} = T_0{}^{**} U_x{}^M.$$

We shall suppose for simplicity that $T_0{}^{**}$ is one-to-one and has its range dense in $\mathfrak{H}(U^L)$. Whenever this is not the case we need only replace U^M by its restriction to the orthogonal complement of the null space of $T_0{}^{**}$ and U^L by its restriction to the closure of the range of $T_0{}^{**}$ and apply the following argument to these two new representations. Making this assumption then about $T_0{}^{**}$, we may write $T_0{}^{**} = WH$, where W is a unitary map of $\mathfrak{H}(U^M)$ on $\mathfrak{H}(U^L)$ and H is non-negative and self-adjoint. Thus $T_0{}^* = HW^*$, so that $T_0{}^* T_0{}^{**} = H^2$. Multiplying together equations (*) and (**) we find that $T_0{}^* T_0{}^{**} = \overset{M}{U}_{x^{-1}} T_0{}^* T_0{}^{**} U_x{}^M$. Thus H^2 and hence H commutes with all $U_x{}^M$. Substituting $T_0{}^{**} = WH$ in (**) and using the fact that H commutes with all $U_x{}^M$ we find that $(U_x{}^L W - W U_x{}^M) H = 0$ and hence that $U_x{}^L W = W U_x{}^M$ for all x. It follows at once that if $K = (1 + H)^{-1}$

then WK^n is an intertwining operator for U^L and U^M for all $n = 0, 1, 2, \cdots$. Since, by hypothesis, $\mathfrak{I}(U^L, U^M)$ is finite, only a finite number of the WK^n can be linearly independent. It follows that the spectrum of K is finite and hence that H must be bounded. Thus T_0^{**} and hence T_0 is bounded. Finally then T_0 has a unique bounded extension to all of $\mathfrak{H}(U^M)$ and this extension is an intertwining operator. Lemma B is thus established.

LEMMA C. *Let $\tilde{\mathfrak{G}}$ be the subgroup of $\mathfrak{G} \times \mathfrak{G}$ consisting of all x, y for which $x = y$. For each double coset $D = (G_1 \times G_2)(x, y)\tilde{\mathfrak{G}}$ let d_D be the dimension of the set of all functions A which satisfy condition* (a), (b) *and* (c) *of Lemma* A *and which vanish outside of D. Then $d = \Sigma_{D \varepsilon \mathfrak{D}} d_D$, where $\mathfrak{D}$ is the family of all $(G_1 \times G_2) : \mathfrak{G}$ double cosets.*

Proof. It is obvious that a function A which satisfies (a) is uniquely determined throughout $(G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ by its value A_{x_0, y_0} at x_0, y_0 and it is also obvious that an A which satisfies (b) and (c) as well will continue to do so after having been reduced to zero outside of this double coset.

LEMMA D. *Let d'_D be the dimension of the set of all functions A which satisfy condition* (a) *of Lemma* A *and which vanish outside of the double coset D. Let x, y be any point of D. Then d'_D is equal to the intertwining number of the representations $s \to L_{xsx^{-1}}$ of the subgroup intertwining number of the representations $s \to L_{xsx^{-1}}$ and $s \to M_{ysy^{-1}}$ of the subgroup $x^{-1}G_1x \cap y^{-1}G_2y$ of $\mathfrak{G}$.*

Proof. A is completely determined by its value at x, y. If this value is B then $A_{\xi xs, \eta ys}$ will be equal to $L_\xi B M_\eta^*$ for all $\xi \varepsilon G_1$, $\eta \varepsilon G_2$ and $s \varepsilon \mathfrak{G}$. Conversely, given a B it will be $A_{x,y}$ for some A if and only if the expression $L_\xi B M_\eta^*$ depends only upon ξxs and ηys and not upon ξ and η; that is, if and only if $\xi_1 x s_1 = \xi_2 x s_2$ and $\eta_1 y s_1 = \eta_2 y s_2$ imply $L_{\xi_1} B M_{\eta_1}^* = L_{\xi_2} B M_{\eta_2}^*$; that is, if and only if $\xi_2^{-1}\xi_1 x = x s_2 s_1^{-1}$ and $\eta_2^{-1}\eta_1 y = y s_2 s_1^{-1}$ imply $L_{\xi_2^{-1}\xi_1} B = B M_{\eta_2^{-1}\eta_1}$; that is, if and only if $x^{-1}\xi x = s$ and $y^{-1}\eta y = s$ imply $L_\xi B = B M_\eta$; that is, if and only if $s \varepsilon x^{-1}G_1x \cap y^{-1}G_2y$ implies $L_{xsx^{-1}}B = BM_{ysy^{-1}}$; that is, if and only if B is an intertwining operator for the representations of the lemma.

LEMMA E. *Let L and M be one-dimensional and let A be a function satisfying* (a), (b) *and* (c) *of Lemma* A. *Then if $A_{x_0, y_0} \neq 0$ the index of $x_0^{-1}G_1x_0 \cap y_0^{-1}G_2y_0$ is finite in both of the containing subgroups $x_0^{-1}G_1x_0$ and $y_0^{-1}G_2y_0$. Conversely, if these two indices are finite then every A which satisfies* (a) *of Lemma* A *and vanishes outside of $(G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ also satisfies* (b) *and* (c). *This last assertion holds without restriction on the dimensions of L and M.*

Proof. If L and M are one-dimensional then A_{x_0,y_0} is a number which is constant in absolute value throughout the double coset to which x_0, y_0 belongs. Now $\| A_{x_0,y_0}(v) \|^2 / \| v \|^2$ is simply the square of this number. Thus condition (b) can be fulfilled only if there are only finitely many summands; that is, only finitely many right-G_1-cosets G_1x for which $x, y_0 \varepsilon (G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$. Similarly condition (c) can be satisfied only if there are only finitely many right-G_2-cosets G_2y for which $x_0, y \varepsilon (G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$. But $x, y_0 \varepsilon (G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ if and only if there exists $\xi \varepsilon G_1$, $\eta \varepsilon G_2$ and $s \varepsilon \mathfrak{G}$ such that $x = \xi x_0 s$ and $y_0 = \eta y_0 s$; that is, if and only if $\xi \varepsilon G_1$ and $\eta \varepsilon G_2$ exist so that $x^{-1}\xi x_0 = y_0^{-1}\eta y_0$, or $x = \xi x_0 y_0^{-1}\eta^{-1} y_0$; that is, if and only if $x \varepsilon (G_1 x_0)(y_0^{-1} G_2 y_0)$. Thus the number of right-G_1-cosets G_1x for which $x, y \varepsilon (G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ is equal to the number of right-G_1-cosets in $(G_1x_0)(y_0^{-1}G_2y_0)$. But $\xi_1 x_0 y_0^{-1}\eta_1 y_0$ and $\xi_2 x_0 y_0^{-1}\eta_2 y_0$ belong to the same G_1-right-coset if and only if ξ exists in G_1 with $\xi\xi_1 x_0 y_0^{-1}\eta_1 y_0 = \xi_2 x_0 y_0^{-1}\eta_2 y_0$; that is, $x_0^{-1}\xi_2^{-1}\xi\xi_1 x_0 = y_0^{-1}\eta_2\eta_1^{-1}y_0$; that is, $y_0^{-1}\eta_2\eta_1^{-1}y_0 \varepsilon x_0^{-1}G_1x_0$. Thus the number of right-G_1-cosets in $(G_1x_0)(y_0^{-1}G_2y_0)$ is equal to the index of $y_0^{-1}G_2y_0 \cap x_0^{-1}G_1x_0$ in $y_0^{-1}G_2y_0$, and if condition (b) is satisfied this index must be finite. Similarly if condition (c) is satisfied the same argument shows that the index of this group in $x_0^{-1}G_1x_0$ must be finite. Conversely, if these two indices are finite then for arbitrary L and M there are only a finite number of summands in (b) and (c). Since these numbers are independent of v and y the desired inequalities are evidently satisfied. This completes the proof of the lemma.

As the reader may easily prove for himself the two indices under discussion in the preceding lemma depend only upon the double coset $(G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ to which x_0, y_0 belongs. Moreover x_0, y_0 and x_1, y_1 belong to the same $(G_1 \times G_2) : \tilde{\mathfrak{G}}$ double coset if and only if $x_0y_0^{-1}$ and $x_1y_1^{-1}$ belong to the same $G_1 : G_2$ double coset. Combining these observations with Lemmas A, B, C, D and E we obtain at once

THEOREM 3′. *Let G_1 and G_2 be both open and closed subgroups of the separable locally compact group $\mathfrak{G}$. Let L and M be unitary representations of G_1 and G_2 respectively. For each x and y in $\mathfrak{G}$ consider the representations $s \to L_{xsx^{-1}}$, $s \to M_{ysy^{-1}}$ of $x^{-1}G_1x \cap y^{-1}G_2y$ and let $\mathfrak{J}(L, M, x, y)$ denote their intertwining number. Then $\mathfrak{J}(M, L, x, y)$ depends only upon the double coset $D = D(x, y) = G_1xy^{-1}G_2$ to which xy^{-1} belongs, so that we may write $\mathfrak{J}(L, M, D)$. Moreover the indices of $x^{-1}G_1x \cap y^{-1}G_2y$ in $x^{-1}G_1x$ and in $y^{-1}G_2y$ depend only upon this double coset. Let $\mathfrak{D}_f$ be the set of all double*

cosets for which both indices are finite and let $\mathfrak{D}$ *be the set of all double cosets. Then*

$$\Sigma_{D\varepsilon\mathfrak{D}_f}\mathfrak{J}(L,M,D) \leq \mathfrak{J}(U^L,U^M) \leq \Sigma_{D\varepsilon\mathfrak{D}}\,\mathfrak{J}(L,M,D);$$

and if L *and* M *are one-dimensional then*

$$\Sigma_{D\varepsilon\mathfrak{D}_f}\mathfrak{J}(L,M,D) = \mathfrak{J}(U^L,U^M).$$

3. Applications. Theorems 4′, 6′ and 7′ are deducible from Theorem 3′ just as were Theorems 4, 6, and 7 from Theorem 3. We leave details to the reader.

THEOREM 4′. *Let* G_1 *be a closed and open subgroup of the separable locally compact group* $\mathfrak{G}$. *Let* L *and* M *be irreducible unitary representations of* G_1 *and* $\mathfrak{G}$ *respectively. Then the number of times that* U^L *contains* M *as a discrete direct summand is less than or equal to the number of times that the restriction of* M *to* G_1 *contains* L *as a discrete direct summand. If* $\mathfrak{G}/G_1$ *is finite then this inequality may be replaced by an equality.*

THEOREM 6′. *Let* $\mathfrak{G}$, G_1 *and* L *be as in Theorem 4′ and let* L *be one-dimensional. Then* U^L *is irreducible if and only if for each* x *not in* G_1 *one of the following statements is true.*

(a) L_s *and* $L_{xsx^{-1}}$ *are not identical for* s *in* $x^{-1}G_1x \cap G_1$.

(b) $x^{-1}G_1x \cap G_1$ *has infinite index in at least one of the two intersecting subgroups.*

We remark here that condition (b) is slightly misstated in [6].

THEOREM 7′ *Let* $\mathfrak{G}$, G_1, G_2, L *and* M *be as in Theorem 3′. Let* L *and* M *be one-dimensional and let* U^L *and* U^M *be irreducible. Then* U^L *and* U^M *are unitary equivalent if and only if there exists* $x \varepsilon \mathfrak{G}$ *such that*

(a) M_s *and* $L_{xsx^{-1}}$ *are identical for* s *in* $x^{-1}G_1x \cap G_2$ *and*

(b) $x^{-1}G_1x \cap G_2$ *has finite index in each intersecting subgroup.*

When L is the identity representation of G_1 then condition (a) of Theorem 6′ is never satisfied. Since condition (b) is independent of L we deduce at once

THEOREM 9. *Let* $\mathfrak{G}$ *and* G_1 *be as in Theorem 6′. A necessary and sufficient condition that* U^L *be irreducible for all one-dimensional representations* L *of* G_1 *is that condition* (b) *of Theorem 6′ hold.*

Theorem 9 in the slightly more special case for which $\mathfrak{G}$ is discrete and G_1 Abelian is due to Godement [4].

We conclude this section with a brief discussion of a class of examples. Let G_1 and G_2 be infinite discrete Abelian groups. Let there be given a homomorphism of G_2 into the group of automorphisms of G_1 and let $x \to y[x]$ denote the automorphism of G_1 defined by $y \in G_2$. Let $\mathfrak{G}$ be the set of all pairs x, y with $x \in G_1$, $y \in G_2$ and let us convert $\mathfrak{G}$ into a group by defining multiplication as follows: $(x_1, y_1)(x_2, y_2) = (x_1 y_1[x_2], y_1 y_2)$. Then the set of all x, e where e is the identity of G_2 is a group isomorphic with G_1 and may be identified with G_1. Similarly the set of all e, y may be identified with G_2. $\mathfrak{G}$ is a "semi-direct product" of G_1 and G_2.

LEMMA 1. *Let G_1 and G_2 be such that all non-trivial orbits of G_1 under G_2 are infinite. Then for each one-dimensional representation M of G_2 the representation U^M is irreducible. Moreover U^M and U^L are unitary equivalent if and only if M and L are the same one-dimensional representation of G_2.*

Proof. A straightforward computation shows that $(x, e)^{-1} G_2 (x, e)$ is the set of all $(x^{-1} y[x], y)$ for $y \in G_2$. Thus $(x, e)^{-1} G_2 (x, e) \cap G_2$ is the set of all $y \in G_2$ with $y[x] = x$. Its index in G_2 is thus the number of distinct transforms of x by elements of y. By hypothesis, then, this index is infinite for $x \neq e$. Since x, e and x, y are in the same $G_2 : G_2$ double coset we conclude that condition (b) of Theorem 6′ holds. The lemma is now seen to be a consequence of Theorems 6′ and 7′.

LEMMA 2. *If L and M are one-dimensional representations of G_1 and G_2 respectively then $\mathfrak{I}(U^L, U^M) = 0$.*

Proof. Since $G_1 G_2 = \mathfrak{G}$ there is only one double coset and it is clearly not in $\mathfrak{D}_f$.

LEMMA 3. *If L is a one-dimensional representation of G_1 then U^L is irreducible if and only if $L(y[x]) \neq L(x)$ for all y in G_2 distinct from the identity.*

Proof. The truth of the lemma is an immediate consequence of Theorem 6′.

It follows from a theorem in [4] (stated more generally in [5]) that the regular representation of $\mathfrak{G}$ is a direct integral over the character group $\hat{G}_1$ of G_1 of the representations U^L for $L \in \hat{G}_1$ and also is a direct integral over the character group $\hat{G}_2$ of G_2 of the representations U^M for $M \in \hat{G}_2$.

Let us now choose as G_1 the additive group of rational numbers and as G_2 the multiplicative group of non-zero rational numbers and let $y[x]$ be the product of x and y. Applying Lemmas 1, 2, and 3 we may conclude that for almost all $L \varepsilon \hat{G}_1$, U^L is irreducible, that for all $M \varepsilon \hat{G}_2$, U^M is irreducible, and that no U^M is unitary equivalent to any U^L or any other U^M. Thus we have

THEOREM 10. *There exists an infinite discrete group whose regular representation may be decomposed into infinite-dimensional irreducible parts in such a manner that no two are unitary equivalent.*

THEOREM 11. *There exists an infinite discrete group whose regular representation may be decomposed into infinite-dimensional irreducible parts in two*[3] *entirely different ways; specifically, so that no component in one decomposition is unitary equivalent to any component in the other.*

4. Concluding remarks. Theorem 3′ is evidently quite far from being a best possible result. The following questions immediately present themselves. (1) Can the inequality be replaced by an equality when L and M are not one-dimensional? (2) Can a result like Theorem 3′ be proved when G_1 and G_2 are not assumed to be open? (3) Can a version of Theorem 3′ be given which uses a less crude definition of intertwining number—one which takes cognizance of the fact that a representation may decompose as a direct integral of irreducible representations rather than as a discrete direct sum of such. Questions (2) and (3) are now under investigation and the results will be reported upon at a later time. Question (1) we have considered only to the extent of observing that the final statement of Theorem 3′ is certainly false unless some restriction is put on L and M. Indeed if in the example just considered we let L and M be the regular representations of G_1 and G_2 we find that $\mathfrak{D}_f$ is empty while $\mathfrak{I}(U^L, U^M) = \infty$.

HARVARD UNIVERSITY.

Added in proof. An independent example of the phenomenon described in Theorem 11 has been found by H. Yoshizawa, "Some remarks on unitary representations of the free group," *Osaka Mathematical Journal,* vol. 3 (1951), pp. 56-63.

[3] For a special class of semi-direct products one of these two decompsitions has been considered by F. I. Mautner [8].

BIBLIOGRAPHY.

[1] Artin, E., "Über eine neue Art von *L*-Reihen," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 3 (1924), pp. 89-108.

[2] Brauer, R., "On Artin's *L*-series with general group characters," *Annals of Mathematics*, vol. 48 (1947), pp. 502-514.

[3] Frobenius, G., "Über Relationen zwischen den Characteren einer Gruppe und denen ihrer Untergruppen," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1898, pp. 501-515.

[4] Godement, R., "Sur la transformation de Fourier dans les groupes discrets," *Comptes Rendus de l'Académie des Sciences (Paris)*, vol. 228 (1949), pp. 627-628.

[5] Mackey, G. W., "Imprimitivity for representations of locally compact groups I," *Proceedings of the National Academy of Sciences, U. S. A.*, vol. 35 (1949), pp. 537-545.

[6] ———, "Imprimitivité pour les représentations des groupes localement compacts. II. Nombres d'entrelacement pour les représentations imprimitives," *Comptes Rendus de l'Académie des Sciences (Paris)*, vol. 230 (1950), pp. 808-809.

[7] ———, "Imprimivité pour les représentations des groupes localement compacts. III. Produits de Kronecker et nombres d'entralacement forts," *Comptes Rendus de l'Académie des Sciences (Paris)*, vol. 230 (1950), pp. 908-909.

[8] Mautner, F. I., "Completeness relations on locally compact groups," *Bulletin of the American Mathematical Society*, vol. 55 (1949), p. 52.

[9] Osima, M., "Note on the Kronecker product of representaitons of a group," *Proceedings of the Imperial Academy of Tokyo*, vol. 17 (1941), pp. 411-413.

[10] Shoda, K., "Über die monomialen Darstellungen einer endlichen Gruppe," *Proceedings of the Physico-Mathematical Society of Japan*, vol. 15 (1933), pp. 249-257.

THE RECIPROCITY FORMULA FOR DEDEKIND SUMS.*

By L. J. Mordell.

Let p, q be two positive integers without a common divisor. It is well known that

$$\sum_{x=1}^{p-1}[qx/p]+\sum_{y=1}^{q-1}[py/q]=(p-1)(q-1), \tag{1}$$

where $[z]$ denotes the integer part of z. There exists another result of this type discovered by Dedekind in discussing the linear transformation of the modular function $\log\eta(\omega)$, one form of which is

$$q\sum_{x=1}^{p-1}x[qx/p]+p\sum_{y=1}^{q-1}y[py/q]=\tfrac{1}{12}(p-1)(q-1)(8pq-p-q-1). \tag{2}$$

Rademacher [2] made a detailed study of this result and has published some five proofs. One is a joint proof with Whiteman, and the last has only just appeared. Some of them are arithmetical in character and quite simple. Another proof has just been given by Rédei [3], and a generalization by Apostol [1]. I notice, however, an entirely different way of considering the subject which is no less simple and relates the result to more general and obvious ones.

Let us consider the sum

$$S=\sum_K(qx+py) \tag{3}$$

extended over the integer sets (x,y) or say the lattice points P lying in the region K defined by

$$0<x<p,\quad 0<y<q,\quad qx+py<pq,$$

and so if O, A, B are the points $(0,0)$, $(p,0)$, $(0,q)$ respectively, K is the open triangle OAB. We call K' the open triangle ACB where C is the point (p,q). We have a 1-1 correspondence between the lattice points $P(x,y)$ in K and $P'(x',y')$ in K' given by

$$x+x'=p,\qquad y+y'=q.$$

Since K and K' together contain $(p-1)(q-1)$ lattice points, K contains exactly $\frac{1}{2}(p-1)(q-1)$ lattice points.

* Received December 5, 1950.

It is well known and obvious that the formula (1) states that the number of lattice points in $K + K'$ is the sum of those in the open triangles OAC and OBC.

In (3), we note that

$$\sum_K x = \sum_{x=1}^{p-1} x[q - qx/p] = \sum_{x=1}^{p-1} (p - x)[qx/p]$$

since $[q - qx/p]$ values of y correspond to given x. Hence

$$\begin{aligned} S &= q\sum_{x=1}^{p-1} (p-x)[qx/p] + p\sum_{y=1}^{q-1} (q-y)[py/q] \\ &= pq\Big(\sum_{x=1}^{p-1} [qx/p] + \sum_{y=1}^{q-1} [py/q]\Big) - \sum_{x=1}^{p-1} qx[qx/p] - \sum_{y=1}^{q-1} py[py/q]. \end{aligned}$$

Then from (1) and (2), we have

$$S = pq(p-1)(q-1) - \tfrac{2}{3}pq(p-1)(q-1) + \tfrac{1}{12}(p-1)(q-1)(p+q+1),$$

and so

$$(4) \qquad S = \tfrac{1}{3}pq(p-1)(q-1) + \tfrac{1}{12}(p-1)(q-1)(p+q+1).$$

Thus the proof of (2) is reduced to the evaluation of the sum (3) as given in (4).

We solve now the more general problem of evaluating

$$(5) \qquad T = \sum_K f(qx + py)$$

where f is an arbitrary polynomial and the summation is extended over the lattice points in K.

Write $\xi = qx + py$ for lattice points x, y in K so that ξ is not divisible by p or q and $0 < \xi < pq$. Then the numbers $pq - \xi$ cannot be represented in this way. For

$$\text{if } pq - \xi = qX + pY, \text{ then } pq = q(x + X) + p(y + Y),$$

and so $x + X \equiv 0 \pmod{p}$. Then $x + X = p$ and similarly $y + Y = q$, and this is clearly impossible. The number of ξ is $\frac{1}{2}(p-1)(q-1)$ and so the ξ and $pq - \xi$ are $(p-1)(q-1)$ in number, and so together they are precisely the integers X not divisible by p or q in the interval $0 < X < pq$.

There is, however, such a representation for the numbers $2pq - \xi$. Write $\xi' = qx' + py'$ for lattice points x', y' in K' so that $pq < \xi' < 2pq$. We have now a 1-1 correspondence between the representation of ξ and ξ' given by

$$x' + x = p, \qquad y' + y = q, \qquad \xi' + \xi = 2pq.$$

We prove two fundamental formulae. The first is

$$(7) \qquad \sum_{y=1}^{q-1} \sum_{x=1}^{p-1} f(qx+py) = \sum_K f(\xi) + \sum_K f(2pq-\xi).$$

The left hand side consists of two parts corresponding to $qx+py<pq$, and to $pq<qx+py<2pq$. The first is $\sum_K f(\xi)$, and on putting $x=p-x'$, $y=q-y'$, the second is $\sum_K f(2pq-\xi)$.

We prove next that

$$(8) \qquad \sum_{X=1}^{pq-1} f(X) - \sum_{X=1}^{p-1} f(qX) - \sum_{X=1}^{q-1} f(pX) = \sum_K f(\xi) + \sum_K f(pq-\xi).$$

The left hand side of this is $\sum f(Y)$ extended over the numbers Y in $0<Y<pq-1$ and not divisible by p or q. These numbers Y can be written as ξ or $pq-\xi$ since the numbers Y in $0<Y<pq-1$ which cannot be represented by ξ are given by $pq-\xi$. This gives (8).

The two equations (7), (8) determine $\sum_K f(\xi)$ for any polynomial $f(\xi)$. Take $f(\xi)=\xi^2$, then (7) becomes

$$\sum_{y=1}^{q-1} \sum_{x=1}^{p-1} (qx+py)^2 = \sum_K \xi^2 + \sum_K (2pq-\xi)^2$$

and so

$$\begin{aligned} &\tfrac{1}{6}q^2(q-1)p(p-1)(2p-1) + \tfrac{1}{6}p^2(p-1)q(q-1)(2q-1) \\ &\qquad + \tfrac{1}{2}p^2q^2(p-1)(q-1) \\ &\qquad = 2\sum_K \xi^2 - 4pq\sum_K \xi + 2p^2q^2(p-1)(q-1), \end{aligned}$$

since K contains $\frac{1}{2}(p-1)(q-1)$ lattice points.

Hence

$$\begin{aligned} (9) \qquad &\tfrac{1}{6}q(q-1)(p-1)(2p-1) + \tfrac{1}{6}p(p-1)(q-1)(2q-1) \\ &\qquad - \tfrac{3}{2}pq(p-1)(q-1) = 2/(pq)\sum_K \xi^2 - 4\sum_K \xi. \end{aligned}$$

Next (8) becomes

$$\sum_{X=1}^{pq-1} X^2 - q^2\sum_{X=1}^{p-1} X^2 - p^2\sum_{X=1}^{q-1} X^2 = \sum_K \xi^2 + \sum_K (pq-\xi)^2$$

or

$$\begin{aligned} &\tfrac{1}{6}pq(pq-1)(2pq-1) - \tfrac{1}{6}q^2p(p-1)(2p-1) \\ &\qquad - \tfrac{1}{6}p^2q(q-1)(2q-1) \\ &\qquad = 2\sum_K \xi^2 - 2pq\sum_K \xi + \tfrac{1}{2}p^2q^2(p-1)(q-1). \end{aligned}$$

This becomes

$$(10)\quad \tfrac{1}{6}(pq-1)(2pq-1)-\tfrac{1}{6}q(p-1)(2p-1)-\tfrac{1}{6}p(q-1)(2q-1)$$
$$-\tfrac{1}{2}pq(p-1)(q-1)$$
$$=2/(pq)\sum_K \xi^2-2\sum_K \xi.$$

These two equations (9), (10) determine $\sum_K \xi$, $\sum_K \xi^2$ and give the result (4). Thus (9) is

$$(11)\quad \tfrac{1}{6}(p-1)(q-1)(-5pq-p-q)=2/(pq)\sum_K \xi^2-4\sum_K \xi,$$

and (10) is

$$(12)\quad \tfrac{1}{6}(p-1)(q-1)(-pq+1)=2/(pq)\sum_K \xi^2-2\sum_K \xi.$$

In fact the left hand side of (10) vanishes when $p=1$ or $q=1$, so we can write it as $\tfrac{1}{6}(p-1)(q-1)(apq+b(p+q)+c)$, where a, b, c are constants.

Equating terms in p^2q^2, $p+q$, 1, clearly

$$\frac{a}{6}=\frac{2}{6}-\frac{1}{2},\ \frac{c}{6}=\frac{1}{6},\ \frac{b}{6}-\frac{c}{6}=-\frac{1}{6},$$

and so $a=-1$, $c=1$, $b=0$.

Hence from (11), (12)

$$(13)\quad 2\sum_K \xi=\tfrac{1}{6}(p-1)(q-1)(4pq+p+q+1)$$

$$(14)\quad 2/(pq)\sum_K \xi^2=\tfrac{1}{6}(p-1)(q-1)(3pq+p+q+2).$$

The result (13) is the required result (4).

It is also clear that on taking $f(\xi)=\xi^{2n}$ in (7), (8), the equations determine $\sum_K \xi^{2n}$, $\sum_K \xi^{2n-1}$ when we know the values of $\sum_K \xi^r$, $0\leqq r\leqq 2n-2$.

If, however, in (7) we replace the function $f(\xi)$ by $f(\xi-pq)$ and subtract from (8), we have

$$(15)\quad \sum_K f(\xi)-\sum_K f(\xi-pq)$$
$$=\sum_{X=1}^{pq-1} f(X)-\sum_{X=1}^{q-1} f(pX)-\sum_{X=1}^{p-1} f(qX)-\sum_{y=1}^{q-1}\sum_{x=1}^{p-1} f(qx+py-pq).$$

Write in the usual notation for the Bernouillian polynomial $B_n(x)$,

$$te^{tx}/(e^t-1)=\sum_{n=0}^{\infty}B_n(x)t^n/n!, \tag{16}$$

so that

$$te^{tx}=\sum_{n=0}^{\infty}(B_n(x+1)-B_n(x))t^n/n!, \qquad B_n(x+1)-B_n(x)=nx^{n-1}.$$

If we take $f(X)=B_n(1+X/(pq))$ in (15), we have the explicit formula for $\sum_K \xi^{n-1}$. This takes the shape

$$n\sum_K \xi^{n-1}=B_n(pq)-p^{n-1}B_n(q)-q^{n-1}B_n(p)+(pq)^{n-1}B_n(1)$$
$$-\sum_{y=1}^{q-1}\sum_{x=1}^{p-1}B_n(x/p+y/q),$$

on noting that

$$\sum_{r=0}^{p-1}B_n(x+r/p)=B_n(px)/p^{n-1}.$$

(as remarked to me by Professor Rademacher).

We can also find an explicit formula for

$$p^{n+1}q^{n+1}\sum_K B_n(x/p+y/q)$$

as a polynomial in p and q. One of a different type has been given by Apostol for odd n by using in a different way lattice points in a triangle.

As well known, $B_n(X)=(-1)^nB_n(1-X)$, from (16). Hence when n is even, we have at once from (8),

$$2\sum_K B_n(x/p+y/q)=\sum_{X=1}^{pq-1}B_n(X/pq)-\sum_{X=1}^{p-1}B_n(X/p)-\sum_{X=1}^{q-1}B_n(X/q).$$

To find the result for all n, write (15) as

$$\sum_K f(\xi/pq)-\sum_K f(\xi/pq-1) \tag{17}$$
$$=\sum_{X=1}^{pq-1}f(X/pq)-\sum_{X=1}^{p-1}f(X/p)-\sum_{X=1}^{q-1}f(X/q)-\sum_{Y=1}^{q-1}\sum_{X=1}^{p-1}f(X/p+Y/q-1).$$

Write $B_n(X)=b_nX^n+b_{n-1}X^{n-1}+\cdots+b_0$. Take

$$f(X-1)=(b_n/\{n+1\})B_{n+1}(X)+\cdots+(b_0/1)B_1(X).$$

Then

$$f(\xi/pq)-f(\xi/pq-1)=b_n/\{n+1\}(B_{n+1}(\xi/pq+1)-B_{n+1}(\xi/pq))$$
$$+\cdots=b_n(\xi/pq)^n+b_{n-1}(\xi/pq)^{n-1}+\cdots=B_n(\xi/pq).$$

Since we can put $\xi = qx + py$, the result is given by (17) on summing for X etc.

UNIVERSITY OF PENNSYLVANIA.

REFERENCES.

[1] T. M. Apostol, "Generalized Dedekind sums and transformation formulae of certain Lambert series," *Duke Mathematical Journal*, vol. 17 (1950), pp. 147-157.

[2] H. Rademacher, "Zur Theorie der Modulfunktionen," *Journal für die reine und angewandte Mathematik*, vol. 167 (1931), pp. 312-336.

———, "Eine arithmetische Summenformel," *Monatshefte für Mathematik und Physik*, vol. 39 (1932), pp. 221-228.

———, "Egy reciprocitásképletről a modulfüggvények elméletéből," (with German résumé), *Matematikai és Fizikai Lapok*, vol. 40 (1933), pp. 24-34.

———, "Die Reziprozitätsformel für Dedekindsche Summen," *Acta Scientiarum Mathematicarum*, vol. 12, part B (1950), pp. 57-60.

——— and A. Whitman, "Theorems on Dedekind sums," *American Journal of Mathematics*, vol. 63 (1941), pp. 377-407.

[3] L. Rédei, "Elementarer Beweis und Verallgemeinerung einer Reziprozitätsformel von Dedekind," *Acta Scientiarum Mathematicarum*, vol. 12, part B (1950), pp. 236-239.

THE EVALUATION OF THE CONSTANT IN THE FORMULA FOR THE NUMBER OF PARTITIONS OF n.*

By D. J. Newman.

In 1918 Hardy and Ramanujan,[1] using the powerful circle method, proved that the number $p(n)$ of partitions of n satisfies

$$p(n) \sim an^{-1} \exp(\pi(2n/3)^{\frac{1}{2}}), \tag{1}$$

where

$$a = 1/(4.3^{\frac{1}{2}}). \tag{2}$$

In 1942 Erdös [2] proved (1), but not (2), by completely elementary methods. The purpose of this paper is to provide an elementary proof of (2), assuming the truth of (1); this together with Erdös' proof gives the full Hardy-Ramanujan theorem without recourse to complex variable methods.

We shall begin by proving

$$\sum_{n=1}^{\infty} n^{-1} \exp(\pi(2n/3)^{\frac{1}{2}} x^n \sim 2(6/\pi)^{\frac{1}{2}} \phi(x) \tag{3}$$

and

$$\sum_{n=1}^{\infty} p(n) x^n \sim (2\pi)^{-\frac{1}{2}} \phi(x), \tag{4}$$

where

$$\phi(x) = (1-x)^{\frac{1}{2}} \exp\left(-\frac{\pi^2}{12} + \frac{\pi^2}{6(1-x)}\right),$$

and the asymptotic relations are meant to hold as x tends to 1 from below. From (1), (3), (4),

$$(2\pi)^{-\frac{1}{2}}\phi(x) \sim \sum p(n)x^n \sim \sum an^{-1}\exp(\pi(2n/3)^{\frac{1}{2}})x^n \sim 2(6/\pi)^{\frac{1}{2}} a\phi(x).$$

A comparison of the first and last of these four asymptotically equivalent functions leads at once to (2).

If we write $x = e^{-u}$ and stipulate that in all subsequent asymptotic

* Received March 8, 1950; revised November 29, 1950.

[1] G. H. Hardy and S. Ramanujan, "Asymptotic formulae in combinatory analysis," *Proceedings of the London Mathematical Society*, ser. 2, vol. 17 (1918), pp. 75-115.

[2] P. Erdös, "On an elementary proof of some asymptotic formulas in the theory of partitions," *Annals of Mathematics*, vol. 43 (1942), pp. 437-450.

formulae x tends to unity from below (*i. e.*, u tends to zero from above), then since $u = \log(1/x) \sim 1 - x$, we have $(1-x)^{\frac{1}{2}} \sim u^{\frac{1}{2}}$. Moreover,

$$\frac{1}{1-x} - \frac{1}{2} \sim \left(\sum_{n=1}^{\infty} n^{-1}(1-x)^n \right)^{-1} = \left(\log \frac{1}{1-(1-x)} \right)^{-1} = u^{-1},$$

so that $\phi(x) \sim u^{\frac{1}{2}} \exp(\pi^2/6u) = \Phi(u)$, say. Hence (3) and (4) may be written

$$(5) \qquad \sum_{n=1}^{\infty} n^{-1} \exp(\pi(2n/3)^{\frac{1}{2}} - nu) \sim 2(6/\pi)^{\frac{1}{2}} \Phi(u),$$

$$(6) \qquad \sum_{n=1}^{\infty} p(n) e^{-nu} \sim (2\pi)^{-\frac{1}{2}} \Phi(u),$$

and it remains only to prove (5) and (6).

We will make use of the following simple result.

LEMMA.[3] *If $f(t)$ is of bounded variation on the positive real axis, then for any $u > 0$,*

$$\left| u \sum_{n=1}^{\infty} f(nu) - \int_0^{\infty} f(t)\, dt \right| \leqq uV,$$

where V is the total variation of f.

Proof of (5). The sum on the left side of (5) is clearly asymptotic to

$$2 \sum_{n=1}^{\infty} n^{-1} e^{-nu} [\sin h(\pi(2n/3)^{\frac{1}{2}}) - \pi(2n/3)^{\frac{1}{2}}]$$

$$= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} 2n^{-1} e^{-nu} (2\pi^2 n/3)^{m+\frac{1}{2}}/(2m+1)!$$

$$= 2 \sum_{m=1}^{\infty} (u^{-m-\frac{1}{2}} (2\pi^2/3)^{m+\frac{1}{2}}/(2m+1)!) (u \sum_{n=1}^{\infty} (un)^{m-\frac{1}{2}} e^{-nu})$$

$$= 2 \sum_{m=1}^{\infty} (u^{-m-\frac{1}{2}} (2\pi^2/3)^{m+\frac{1}{2}}/(2m+1)!) \Gamma(m+\tfrac{1}{2}) (1 + O(u)),$$

by the lemma cited above and the fact that the total variation of $t^{m-\frac{1}{2}} e^{-t}$ is

$$2(m-\tfrac{1}{2})^{m-\frac{1}{2}} e^{-m+\frac{1}{2}} = O(1)\Gamma(m+\tfrac{1}{2}).$$

Since

$$\Gamma(m+\tfrac{1}{2})/(2m+1)! = \pi^{\frac{1}{2}} 2^{-2m-1}/m!(m+\tfrac{1}{2}),$$

---

[3] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 1 (1925), p. 37, where it is proved for the interval $(0, 1)$. It is easily extended to the interval $(0, \infty)$.

it follows that the sum on the left side of (5) is asymptotic to

$$2(6u/\pi)^{\frac{1}{2}}\sum_{m=1}^{\infty}(\pi^2/6u)^{m+1}/m!(m+\tfrac{1}{2})$$

$$=2(6u/\pi)^{\frac{1}{2}}\sum_{m=1}^{\infty}((\pi^2/6u)^{m+1}/(m+1)!)(1+O(1/m))$$

$$\sim 2(6u/\pi)^{\frac{1}{2}}\exp(\pi^2/6u).$$

Proof of (6). Since $\sum_{n=1}^{\infty} p(n)x^n = \sum_{m=1}^{\infty}(1-x^m)^{-1}$, we have

$$\log(\sum_{n=1}^{\infty} p(n)e^{-nu}) = \sum_{n=1}^{\infty} n^{-1}(e^{nu}-1)^{-1}$$

$$= -\sum_{n=1}^{\infty}(n^{-2}u^{-1}-(2n)^{-1}e^{-nu}-n^{-1}(e^{nu}-1)^{-1}) + \pi^2/6u + 2^{-1}\log(1-e^{-u}),$$

and the last term may be replaced by $2^{-1}\log u + o(1)$.

Once again we may apply the lemma. It is well-known [4] that

$$\int_0^{\infty}(t^{-2}-(2t)^{-1}e^{-t}e^{-t}-t^{-1}(e^t-1)^{-1})\,dt = 2^{-1}\log 2\pi.$$

Hence

$$\log(\sum_{n=1}^{\infty} p(n)e^{-nu}) = 2^{-1}\log(u/2\pi) + \pi^2/6u + o(1),$$

which implies the desired result.

NEW YORK UNIVERSITY.

[4] E. T. Whittaker and G. N. Watson, *Modern Analysis*, American Edition, New York, 1944. On p. 249, line 15, let z tend to zero and apply p. 248, line 12.

UNIFORM CAUCHY POINTS AND POINTS OF EQUICONTINUITY.* [1]

By B. J. Pettis.

Let X and Y be metric spaces with metrics ρ and σ, and let $\{f_n\}$ be a sequence of continuous functions on X to Y converging in X to a function f_0. A point x_0 is *a point of uniform convergence of* $\{f_n\}$ *to* f_0 if for each $\epsilon > 0$ there exist a $\delta_\epsilon > 0$ and a positive integer n_ϵ such that $\sigma(f_n(x), f_0(x)) < \epsilon$ holds whenever $\rho(x, x_0) < \delta_\epsilon$ and $n \geqq n_\epsilon$. Let U be the set of all such x_0. In 1897, antedating Baire's thesis, W. F. Osgood defined a set E in $[0,1]$ to be a "P-set" [15, p. 161] if it is nowhere dense and "contains its derivative," i. e., E is nowhere dense and closed, and called E a "Q-set" [15, p. 171] if it is the union of a monotone increasing sequence of P-sets, that is, if it is a first category F_σ-set. He proved that the complement $X - U$ of U is a Q-set when X is $[0,1]$, Y the reals, and f_0 is continuous. Shortly afterwards other authors [6, pp. 1140-3] removed the continuity hypothesis on f_0 and showed in addition that f_0 is continuous at each point of U. In 1924 Kuratowski [10] extended this to the case in which X is separable and Y metric, and in 1930 Banach [1] eliminated Kuratowski's separability assumption. An excellent proof for topological X and metric Y can be found in Hausdorff's book [9, pp. 285-6].

In this note we wish to extend the Osgood theorem to the case in which $\{f_\lambda\}$ is a directed set of continuous functions, X is a topological space, Y is a uniform space, and no limit function f_0 is assumed; in place of the omitted condition it is assumed that the set C of Cauchy points of $\{f_\lambda\}$ is, in a certain prescribed sense, a "residual set" in X. The resulting conclusion is that the set C_u composed of uniform Cauchy points and the set C_a of points of almost equi-continuity also are residual in the same sense. (The case in which f_0 exists is covered in Corollary 1.1; the latter thus constitutes a direct generaliaztion of Hausdorff's version of the Osgood theorem.) The proof of Theorem 1 combines that of Hausdorff with a device due to A. Weil.

* Received June 13, 1950; revised October 20, 1950.

[1] Presented to the American Mathematical Society September 6, 1948, under the title *On points of uniform convergence in topological spaces.* This paper was written under Contract N7-onr-434, Task Order III, Navy Department (the Office of Naval Research).

In Theorem 2 we consider the case in which C has the property that its complement is not residual; the conclusion is that C_u and C_a have the same property. Applying Theorems 1 and 2 to the case in which $\{f_\lambda\}$ is essentially denumerable and Y is pseudo-metric, it follows (Theorem 3) that (i) if the complement of C is first category so are the complements of C_u and C_a, and (ii) if C is second category so are C_u and C_a. Several well known results in the literature can be considered as applications of Theorem 3; some of these—the theorems of Lebesgue-Hahn-Saks and of Nikodym concerning convergent sequences of completely additive set functions, the uniform boundedness of convergent sequences of linear operations in Banach spaces, the theorem of Montgomery on the double continuity of the group product in certain groups with topologies, Yosida's abstraction of Banach's theorem on convergence almost everywhere, etc.—are derived below and in certain cases in more generality.

When $\{f_\lambda\}$ is a denumerable sequence and Y is metric and X is a complete metric space the conclusions of Theorem 3 concerning C are identical with recent results due [2] to A. Alexiewicz [20, p. 5], results from which the latter has drawn a variety of theorems including two of the applications of Theorem 3 that were mentioned above. The two papers differ in that each contains applications not in the other, and in that Theorems 1, 2, and 3 of the present paper constitute generalizations of Alexiewicz's Theorem 1 and Corollary.

Let X be any topological space and ϕ any cardinal number. A subset E of X is a I_ϕ-set if it is the union of ϕ nowhere dense sets, and otherwise is a II_ϕ-set; it is ϕ-residual if its complement $X - E$ is a I_ϕ-set [5]. Let Y be any completely regular topological space, not necessarily Hausdorff, and of all uniformizations of the topology in Y [18] let $\{V_\alpha \mid \alpha \varepsilon A\}$ be one having least cardinal number $|A|$. For each element λ of a fixed directed set Λ [3] let f_λ be a function on X to Y. A point x_0 in X is *a Cauchy point of* $\{f_\lambda\}$ if (0.1) given any $\alpha \varepsilon A$ there is some $\lambda_\alpha \varepsilon \Lambda$ such that $f_\lambda(x_0) \varepsilon V_\alpha(f_{\lambda_\alpha}(x_0))$ for all $\lambda \geqq \lambda_\alpha$; and x_0 is *a uniform Cauchy point* [3] *of* $\{f_\lambda\}$ *if* (0.2) for each $\alpha \varepsilon A$ there exist a $\lambda_\alpha \varepsilon \Lambda$ and an open set G about x_0 such that $f_\lambda(x) \varepsilon V_\alpha(f_{\lambda_\alpha}(x))$ for all $\lambda \geqq \lambda_\alpha$ and all x in G. The set of Cauchy points will be denoted by C and the set of uniform Cauchy points by C_u. *A point*

---

[2] It was after submission of the present paper for publication that the author became aware of [20]. Results common to the two papers are clearly to the credit of Alexiewicz.

[3] A term suggested by the referee, in place of Du Bois-Reymond's *a point of uniform convergence* for a sequence $\{f_n\}$.

of almost equi-continuity of $\{f_\lambda\}$ is a point x_0 satisfying (0.3) for each $\alpha \varepsilon A$ there exist a $\lambda_\alpha \varepsilon \Lambda$ and an open G about x_0 such that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(x_0))$ for any $\lambda \geqq \lambda_\alpha$ and any x in G. The set of all such points will be denoted by C_a. If x_0 has the property that for each α there is an open G about x_0 such that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(x_0))$ for all x in G and all λ, we say that x_0 is *a point of equi-continuity of* $\{f_\lambda\}$, and denote by C_e the set of all such x_0. Clearly $C_e \subset C_a$, and if Λ is the set of positive integers and each f_λ is continuous then $C_e = C_a$. The closure and interior of any set E in a topological space will be represented by E^* and E^0 respectively.

The following lemma, which is well known when Y is metric and Λ is the set of positive integers [8], is easily verified in the present case.

LEMMA 1. *The relations $C_u \supset C^* \cap C_a = C \cap C_a$ always hold, and if f_λ is continuous for each λ in a cofinal subset* [3] *of Λ then $C_u = C^* \cap C_a = C \cap C_a$.*

Letting $|\Lambda|$ represent the least of the cardinal numbers belonging to cofinal subsets of Λ, our main result is

THEOREM 1. *Suppose that each f_λ is continuous, that $\phi \geqq \max(|A|, |\Lambda|)$ and that C is ϕ-residual. Then C_u and C_a are ϕ-residual and $C_u \subset C_a$.*

For each $\alpha \varepsilon A$ there exists [18, p. 14] a non-negative real function $\sigma_\alpha(p, q)$ on $Y \times Y$ such that (i) $\sigma_\alpha(p, p) = 0$, (ii) $\sigma_\alpha(p, q) = \sigma_\alpha(q, p)$, (iii) if $\sigma_\alpha(p, q) < 1$ then $p \varepsilon V_\alpha(q)$ and $q \varepsilon V_\alpha(p)$, and (iv) σ_α is uniformly continuous, i. e., given $\epsilon > 0$ there is some $\beta \varepsilon A$ such that $|\sigma_\alpha(r, s) - \sigma_\alpha(p, q)| < \epsilon$ whenever $r \varepsilon V_\beta(p)$ and $s \varepsilon V_\beta(q)$.

Let $M = [\mu]$ be a cofinal subset of Λ having cardinal number $|\Lambda|$, and for each $\alpha \varepsilon A$ and $\mu \varepsilon M$ let $Q_{\alpha\mu} = X[x \mid |\sigma_\alpha(f_\lambda(x), f_\mu(x))| \leqq \frac{1}{2}$ for all $\lambda \geqq \mu]$ and $R_{\alpha\mu} = X - Q_{\alpha\mu}$. The essential part of the proof is to establish, independent of the continuity of the f_λ's, that

$$C = \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu} \text{ and } C_u = \bigcap_\alpha \bigcup_\mu Q^0{}_{\alpha\mu}.$$

The first of these implies that $X - C = \bigcup_\alpha \bigcap_\mu R_{\alpha\mu}$; since $X - C$ is a I_ϕ-set, so is $\bigcap_\mu R_{\alpha\mu}$ for each α. From the second equality, $X - C_u = \bigcup_\alpha \bigcap_\mu (X - Q^0{}_{\alpha\mu}) = \bigcup_\alpha \bigcap_\mu R^*{}_{\alpha\mu}$; if $\bigcap_\mu R^*{}_{\alpha\mu}$ were a I_ϕ-set for each α then, since $\phi \geqq |A|$, clearly $X - C_u$ must be a I_ϕ-set, i. e., C_u is ϕ-residual, and this combined with Lemma 1 provides the desired conclusion. Now $\bigcap_\mu R^*{}_{\alpha\mu} \subset [\bigcap_\mu R_{\alpha\mu}] \cup [\bigcup_\mu (R^*{}_{\alpha\mu} - R_{\alpha\mu})]$, where $\bigcap_\mu R_{\alpha\mu}$ has been noted to be a I_ϕ-set; since $\phi \geqq |\Lambda|$, $\bigcap_\mu R^*{}_{\alpha\mu}$ will then be a I_ϕ-set if each $R^*{}_{\alpha\mu} - R_{\alpha\mu}$ is a I_ϕ-set. But for fixed λ and μ the function $\sigma_\alpha(f_\lambda(x), f_\mu(x))$ is continuous in x by (iv)

above, since f_λ and f_μ are continuous; thus each $Q_{\alpha\mu}$ is closed, each $R_{\alpha\mu}$ is open, and hence each $R^*_{\alpha\mu} - R_{\alpha\mu}$ is nowhere dense and therefore a I_ϕ-set.

To verify the equality $C = \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu}$, first suppose $x_0 \varepsilon C$ and α is fixed. By (iv) and (i) above there is a $\beta \varepsilon A$ such that $r \varepsilon V_\beta(p)$ and $s \varepsilon V_\beta(p)$ imply $|\sigma_\alpha(r, s)| < \frac{1}{2}$. Having $\beta \varepsilon A$, $x_0 \varepsilon C$, and M cofinal in Λ, there is some $\mu_\beta \varepsilon M$ such that $f_\lambda(x_0) \varepsilon V_\beta(f_{\mu_\beta}(x_0))$ for all $\lambda \geqq \mu_\beta$. Since $f_{\mu_\beta}(x_0) \varepsilon V_\beta(f_{\mu_\beta}(x_0))$, it follows that $|\sigma_\alpha(f_\lambda(x_0), f_{\mu_\beta}(x_0))| < \frac{1}{2}$ for $\lambda \geqq \mu_\beta$, and hence $x_0 \varepsilon Q_{\alpha\mu_\beta}$. Thus $C \subset \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu}$. Conversely, if $x_0 \varepsilon \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu}$ then $x_0 \varepsilon Q_{\alpha\mu_\alpha}$ for each α, i. e., $|\sigma_\alpha(f_\lambda(x_0), f_{\mu_\alpha}(x_0))| \leqq \frac{1}{2}$ for all $\lambda \geqq \mu_\alpha$, so that by (iii) we have $f_\lambda(x_0) \varepsilon V_\alpha(f_{\mu_\alpha}(x_0))$ for all such λ. Thus $x_0 \varepsilon C$.

To establish the second equality, fix x_0. If for each α we have $x_0 \varepsilon Q^0{}_{\alpha\mu_\alpha}$ for some $\mu_\alpha \varepsilon M$ then, letting $G = Q^0{}_{\alpha\mu_\alpha}$, clearly $|\sigma_\alpha(f_\lambda(x), f_{\mu_\alpha}(x))| \leqq \frac{1}{2}$ holds for $\lambda \geqq \mu_\alpha$ and $x \varepsilon G$, so that by (iii) we have $f_\lambda(x) \varepsilon V_\alpha(f_{\mu_\alpha}(x))$ for such λ and x. Since this holds for each α, the point x_0 is in C_u. On the other hand, when x_0 is in C_u consider any fixed α. By (iv) and (i) there is some $\beta \varepsilon A$ such that $|\sigma_\alpha(r, s)| < \frac{1}{2}$ whenever $r \varepsilon V_\beta(p)$ and $s \varepsilon V_\beta(p)$. Having x_0 in C_u and M cofinal in Λ, there is some open G about x_0 and some $\mu_\beta \varepsilon M$ such that $f_\lambda(x) \varepsilon V_\beta(f_{\mu_\beta}(x))$ whenever $x \varepsilon G$ and $\lambda \geqq \mu_\beta$. For such x and λ we then have $|\sigma_\alpha(f_\lambda(x), f_{\mu_\beta}(x))| < \frac{1}{2}$, implying that $G \subset Q_{\alpha\mu_\beta}$ and hence $x_0 \varepsilon Q^0{}_{\alpha\mu_\beta}$, ending the proof.

It may be remarked at this point that the above proof that C_u is ϕ-residual still holds in case there is a secondary topology in X in which every open set has, in terms of the original topology, a I_ϕ-set for its boundary; if under these circumstances each f_λ is assumed continuous in the secondary topology rather than in the primary one the proof goes through with little change, the conclusions still being in terms of the primary topology.

Given any function f_0 on X to Y, let C_0 be the set of points of continuity of f_0; and given f_0 and $\{f_\lambda\}$, let $C(f_0)$ and $C_u(f_0)$ represent the sets of those points x_0 for which (0.1) and (0.2) respectively hold when f_{λ_α} is replaced by f_0. The verification of the following is similar to that of Lemma 1 and is also omitted.

LEMMA 2. *For any* f_0 *and any* $\{f_\lambda\}$ *these hold*: (1) $C \supset C(f_0)$; (2) $C_0 \supset C(f_0)^0 \cap C_a$; (3) $C(f_0) \cap C_u \supset C_u(f_0) \supset C(f_0)^0 \cap C_u$; (4) *if* f_λ *is continuous for each* λ *in a cofinal set then* $C_0 \supset C_u(f_0)$.

The following form of the Osgood theorem can now be derived.

COROLLARY 1. 1. *Given* $\{f_\lambda\}$ *and* f_0, *suppose that each* f_λ *is continuous.*

that $\phi \geqq \max(|A|, |\Lambda|)$, *and that* $C(f_0)^0$ *is* ϕ*-residual. Then* $C_u(f_0)$, $C_a \cap C_0$, *and* C_0 *are* ϕ*-residual, and* $C_u(f_0) \subset C_a \cap C_0$.

If $C(f_0)^0$ is ϕ-residual so is C by (1) of Lemma 2; from Theorem 1, C_u must be ϕ-residual. The intersection $C(f_0)^0 \cap C_u$ then has the same property; and since the lemmas give us $C(f_0)^0 \cap C_u \subset C_u(f_0) \subset C_u \subset C_a$ and $C_u(f_0) \subset C_0$, the corollary follows.

THEOREM 2. *Suppose that each* f_λ *is continuous, that* $\phi \geqq \max(|A|, |\Lambda|)$, *and that* C *is a* II_ϕ*-set. Then* $C \cap C_a$ $(= C_u)$ *is a* II_ϕ*-set.*

Let $g_\lambda(x) = f_\lambda(x)$ for $x \varepsilon C$. By Theorem 1 there is a set D which is ϕ-residual in the space C and has each of its points a point of almost equi-continuity of $\{g_\lambda\}$ in C. From its first property D is a II_ϕ-set in X since C is, and hence [5, Th. 3. 3] there exists a non-null open set G_D in X such that $N \cap G_D \cap D$ is a II_ϕ-set in X whenever N is open in X and $N \cap G_D$ is non-null. Letting $T = G_D \cap D$, T is a II_ϕ-set and $T \subset C$; since $C \cap C_a = C_u$ by Lemma 1, we have only to prove that $T \subset C_a$.

Fix $t \varepsilon T$ and $\alpha \varepsilon A$, and choose β so that if p, q, $r \varepsilon Y$, $p \varepsilon V_\beta(q)$, and $p \varepsilon V_\beta(r)$ then $q \varepsilon V_\alpha(r)$. Now $t \varepsilon D$ implies that for some neighborhood $N(t)$ of t in X and some $\lambda_\beta \varepsilon \Lambda$ we have $g_\lambda(x') \varepsilon V_\beta(g_\lambda(t))$ for any $\lambda \geqq \lambda_\beta$ and any $x' \varepsilon N(t) \cap C$. If it can be established that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(t))$ for any $\lambda \geqq \lambda_\beta$ and any x in the set $N(t) \cap G_D$, which is open in X and contains t, clearly t is in C_a. Fixing x and λ accordingly, the continuity of f_λ means that $f_\lambda(x') \varepsilon V_\beta(f_\lambda(x))$ for any x' in some set $N(x)$ open in X about x. Now $N(x) \cap N(t) \cap G_D$ is open and is non-null since it contains x, and hence $N(x) \cap N(t) \cap T$ is a II_ϕ-set and therefore non-null. Choose any x' in this set. Since $x' \varepsilon N(x)$ we have $f_\lambda(x') \varepsilon V_\beta(f_\lambda(x))$; and since $x' \varepsilon N(t) \cap T \subset N(t) \cap C$, a remark above yields the inclusion $g_\lambda(x') \varepsilon V_\beta(g_\lambda(t))$, whence $f_\lambda(x') \varepsilon V_\beta(f_\lambda(t))$. From the manner in which β was chosen we conclude that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(t))$, completing the proof.

Defining $\{f_\lambda\}$ to be *of Class* 1-ϕ if C_a is either a I_ϕ-set or the whole space X, there is

COROLLARY 2.1. *Suppose in addition to the hypotheses of Theorem 2 that* $\{f_\lambda\}$ *is of Class* 1-ϕ. *Then* $C_a = X$ *and the sets* C_u, C, C^* *coincide and contain an open* II_ϕ*-set.*

From the theorem C_a is a II_ϕ-set and so $C_a = X$; from Lemma 1 we then have $C_u = C^* = C$. Since C is a closed II_ϕ-set its interior C^0 is a II_ϕ-set.

COROLLARY 2.2. *Suppose in addition to the hypotheses of Corollary* 2.1 *that X is a topological group and C is a subgroup of X. Then $C_a = X$, and C is an open and closed II_ϕ-subgroup; if X is connected then $C = C_a = C_u = X$ and if also Y is complete* [18] *then* $\lim_\lambda f_\lambda(x)$ *exists everywhere in X and is continuous.*

Corollary 2.2 is an extension of a theorem of Banach's [2].

Calling Λ *essentially denumerable* if $|\Lambda| \leqq \aleph_0$ (e.g., $\Lambda =$ positive integers, $\Lambda =$ positive reals) and labeling a topological space Y *pseudo-metric* if the topology is given by a function σ having the properties of a metric except possibly that of $\sigma(y_1, y_2)$ being zero implying $y_1 = y_2$, the following is evident, the uniform topology in Y being that determined by σ. This result (in slightly less general form) is due to Alexiewicz [20].

THEOREM 3. *Let Y be pseudo-metric and Λ essentially denumerable and suppose f_λ is continuous for each $\lambda \in \Lambda$. Then* (1) *if $X - C$ is first category so are $X - C_a$ and $X - C_u$;* (2) *if C is second category so are C_a and C_u;* (3) *if f_0 is on X to Y and $X - C(f_0)^0$ is first category so are $X - C_u(f_0)$, $X - C_a$, and $X - C_0$. Hence if X is second category so are C_u and C_a in all three cases, and in the case of* (3) *so is $C_u(f_0) \cap C_a \cap C_0$.*

The last conclusion in Theorem 3 holds when X contains a conditionally compact non-null open set or when X is a locally complete pseudo-metric space.

Remarks on topological groups. The above results, being essentially double limit theorems, may be expected to have applications of the following kinds. The first is an extension of a well known result due to D. Montgomery [13].

THEOREM 4. *Let X be a group and also a space with a uniform structure, and suppose that xy is continuous in each variable separately. Suppose that ϕ is the least cardinal number among those belonging to families of neghborhoods fundamental in the uniform structure. If X is a II_ϕ-set then xy is continuous in (x, y).*

If xy is not continuous at a point (a, b) in $X \times X$ it is not continuous at (e, e) where e is the identity element. Let $\{V_\lambda \mid \lambda \in \Lambda\}$ be a fundamental family of neighborhoods having cardinal number ϕ. Since xy is not continuous at (e, e) there is an open set N_0 containing e and points x_λ, y_λ in each $V_\lambda(e)$ such that $x_\lambda y_\lambda$ is not in N_0. For each λ let $f_\lambda(x) = xy_\lambda$; each f_λ is continuous in x. Partially ordering Λ by defining $\mu \geqq \lambda$ to mean

$V_\mu(e) \subset V_\lambda(e)$, Λ is a directed set and, letting f_0 be the identity map in X, we have $C(f_0) = X$ since xy is continuous in y. Applying Corollary 1.1, $C_u(f)$ is ϕ-residual; since X is a II_ϕ-set $C_u(f_0)$ must be non-null. Choose $x_0 \varepsilon C_u$, pick $\beta \varepsilon \Lambda$ so that $x_0^{-1}V_\beta(x_0) \subset N_0$, and let γ in Λ be such that $p \varepsilon V_\gamma(q)$ and $q \varepsilon V_\gamma(r)$ imply $p \varepsilon V_\beta(r)$. Since x_0 is in $C_u(f_0)$, given γ there are δ and ϵ in Λ such that $xy_\lambda \varepsilon V_\gamma(x)$ whenever $\lambda \geqq \delta$ and $x \varepsilon V_\epsilon(x_0)$. Choosing first $\zeta \geqq \gamma$, ϵ and then η such that $V_\eta(e) \subset x_0^{-1}V_\zeta(x_0)$, clearly $\lambda \geqq \eta$ implies $x_0 x_\lambda \varepsilon V_\zeta(x_0) \subset V_\gamma(x_0) \cap V_\epsilon(x_0)$. Now take any fixed $\mu \geqq \delta, \eta$. Since $\mu \geqq \eta$, it follows that $x_0 x_\mu \varepsilon V_\epsilon(x_0)$ and hence, since $\mu \geqq \delta, \eta$, that $x_0 x_\mu y_\mu \varepsilon V_\gamma(x_0 x_\mu)$. But $\mu \geqq \eta$ also gives us $x_0 x_\mu \varepsilon V_\gamma(x_0)$. These two inclusions imply that $x_0 x_\mu y_\mu \varepsilon V_\beta(x_0)$, so that $x_\mu y_\mu \varepsilon x_0^{-1}V_\beta(x_0) \subset N_0$, which is contrary to the way in which x_μ and y_μ were originally chosen.

(2) For the second application let X be a topological group and let G be the group of all single-valued functions g on X to X with the group operation $(gh)(x) = g(x)h(x)$. Let H be a subgroup of G possessing a pseudo-metric ν with these properties: (i) ν is right invariant, i.e., $\nu(g, h) = \nu(g\bar{h}, h\bar{h})$ for any $g, h, \bar{h}$ in H, (ii) H is second category with respect to ν, (iii) $\lim_n \nu(h_n, h_0) = 0$ implies $\lim_n h_n(x) = h_0(x)$ for every x in X. Suppose also that given any nucleus U (open set about the identity e) in X there is a right invariant pseudo-metric ρ in X, depending on U, such that (iv) each ρ-sphere $S_\rho(a;\epsilon) \equiv X[x \mid \rho(x,a) < \epsilon]$ is open in X, (v) $S_\rho(e;1) \subset U$, and (vi) each h in H is continuous on X to X with respect to the ρ-topology. For one example, let X be a real or complex linear space with a topology such that X is a topological group satisfying the first countability axiom and scalar multiplication hx is continuous in h and x separately; by the Birkhoff-Kakutani metrization theorem there is a right invariant pseudo-metric $\bar{\rho}$ giving the group topology, and if H is taken to be all functions hx, h a scalar, and is given the scalar topology then (i)-(vi) are all satisfied when for a given U we take ρ to be $n\bar{\rho}$ with n chosen so that $S_{\bar{\rho}}(e;1/n) \subset U$. For another, let X be a real linear space with a topology such that X is a topological group, X has arbitrarily small non-null open convex sets about any point, and hx is continuous in each separate variable; let H again be the scalar multiples, and given U choose a convex nucleus K in U and let ρ be the pseudo-metric determined by the Minkowski functional of K. The next theorem, when applied to the first example, yields a result of Mazur and Orlicz [12, p. 188].

THEOREM 5. *Under the above assumptions on X and H the function $h(x)$ on $H \times X$ to X is continuous.*

If it were not continuous there would exist $x_0 \varepsilon X$, $h_0 \varepsilon H$, and an open set G about $h_0(x_0)$ such that for no choice of N open about x_0 and of $\delta > 0$ do we have $h(x) \varepsilon G$ for all x in N and all h with $\nu(h, h_0) < \delta$. Writing $\bar{x}$ for $h_0(x_0)$ and $U\bar{x}$ for G where U is a nucleus in X, choose ρ so that (iv)-(vi) hold. From (v) and the right invariance of ρ we have $S_\rho(\bar{x}; 1) \subset U\bar{x} = G$, and from (iv) $S_\rho(x_0; 1/n)$ is open about x_0 for $n = 1, 2, \cdots$. Hence there exist $x_n \varepsilon X$ and $h_n \varepsilon H$, $n = 1, 2, \cdots$, such that $\rho(x_n, x_0) < 1/n$, $\nu(h_n, h_0) < 1/n$, and $h_n(x_n)$ is not in G, i.e., $\rho(h_n(x_n), h_0(x_0)) \geqq 1$. Now let $f_n(h) = h(x_n)$. Condition (iii) obviously implies that each f_n is continuous on H to X; and for each h we have, by (vi), that $f_n(h) \to h(x_0)$ in the ρ-topology since $\rho(x_n, x_0) \to 0$. Applying (2) of Theorem 3 there exists a point $\bar{h}$ of almost equi-continuity of $\{f_n\}$. Hence there is a $\delta > 0$ such that $\rho(h(x_n), \bar{h}(x_n)) < \frac{1}{2}$ whenever n is sufficiently large and $\nu(h, \bar{h}) < \delta$. Now $\nu(h_n, h_0) < \delta$ for sufficiently large n and hence, using (i), $\nu(h_n h_0^{-1}\bar{h}, \bar{h}) < \delta$; it then follows that for large n we have $\frac{1}{2} > \rho(h_n(x_n) h_0(x_n)^{-1}\bar{h}(x_n), \bar{h}(x_n))$ $= \rho(h_n(x_n), h_0(x_n))$. But $\rho(x_n, x_0) \to 0$ and condition (vi) together imply that $\rho(h_0(x_n), h_0(x_0)) < \frac{1}{2}$ for large n. Hence $1 > \rho(h_n(x_n), h_0(x_0))$ for large n, which contradicts the choice of $\{x_n\}$ and $\{h_n\}$.

(3) In this section *X and Y are topological groups, Λ is essentially denumerable, and $\{h_\lambda \mid \lambda \varepsilon \Lambda\}$ is a family of continuous homomorphisms on X to Y.* The sets C, C_u, C_a, and C_e in X determined by the right uniform topology in Y will be denoted by $C(R)$, $C_u(R)$, etc., and those determined by the left uniform topology by $C(L)$, $C_u(L)$, etc. For any family of homomorphisms the sets $C_a(R)$ and $C_a(L)$ have these properties: (i) each is either the null set or the whole space, (ii) each contains e if and only if the other does. Hence $C_a(R) = C_a(L)$ and either both are null or both are the whole space. (This implies that any such family is of Class 1-ϕ for any ϕ and for any one of the two uniform topologies considered in Y; the next theorem may thus be considered as a variation of Corollary 2.2.)

THEOREM 6. *Suppose that either $C(R)$ or $C(L)$ is second category. Then $C_a(R) = C_a(L) = X$, $C_u(R) = C(R) = C(R)^*$, and $C_u(L) = C(L)$ $= C(L)^*$; if Λ is the positive integers then also $C_e(R) = C_e(L) = X$.*

It is sufficient to show that for any nucleus V in Y a nucleus U can be found such that $h_\lambda(x) \varepsilon V$ holds for $x \varepsilon U$ and for sufficiently large λ. Suppose that $C(R)$ is second category. Given V, let σ be a right-invariant pseudometric such that $S_\sigma(e; 1) \subset V$ and each σ-sphere $S_\sigma(b; \epsilon)$ is open in the Y-topology [16]. The latter fact implies, using the continuity of h_λ, that h_λ is continuous on X to $[Y, \sigma]$. It also implies that for any $\epsilon > 0$ the set

$S_\sigma(e;\epsilon)y$ is one of the uniform neighborhoods in the right uniform topology in Y; hence for any x in C there is some λ_ϵ such that $h_\lambda(x) \varepsilon S_\sigma(e;\epsilon)h_{\lambda_\epsilon}(x)$ for all $\lambda \geqq \lambda_\epsilon$. By the right invariance of σ we have $h_\lambda(x) \varepsilon S_\sigma(h_{\lambda_\epsilon}(x);\epsilon)$ for $\lambda \geqq \lambda_\epsilon$, so that $\{h_\lambda(x)\}$ is a Cauchy sequence in the σ-topology for each $x \varepsilon C$. By Theorem 3 there is a point a in X, a nucleus U, and some $\lambda_1 \varepsilon \Lambda$ such that $\sigma(h_\lambda(x), h_\lambda(a)) < 1$ for all $x \varepsilon Ua$ and all $\lambda \geqq \lambda_1$. Using the right invariance of σ again it follows that $\sigma(h_\lambda(x'), e) < 1$ (and hence $h_\lambda(x') \varepsilon V$) for all $x' \varepsilon U$ and all $\lambda \geqq \lambda_1$. Thus $e \varepsilon C_a(R)$ and hence $C_a(R) = C_a(L) = X$. The rest of the theorem now follows from Lemma 1 and a remark in the introduction. The case in which $C(L)$ is second category follows by a dual argument.

When Λ is the positive integers and X and Y are spaces of Type (F) the theorem just completed was established by Mazur and Orlicz [11]. Together with (2) of Lemma 2, Theorem 6 yields the following, proved for Λ the positive integers and for complete metric X and metric Y by Banach [2, 16].

COROLLARY 6.1. *If $h_\lambda(x)$ converges to $h_0(x)$ for every x in X, where X is second category and Y is a Hausdorff group, h_0 is a continuous homomorphism.*

It is also obvious that when X and Y are linear normed spaces and Λ is the set of positive integers Theorem 6 reduces to the well known assertion that $\{h_\lambda\}$ is uniformly bounded.

(4) For the final application to groups let X and Y be additively written topological groups, with θ denoting the zero element in each. Suppose also that in Y there are defined a partial ordering and a pseudo-metric σ such that (i) σ gives the topology in Y, (ii) for any sequence $\{y_n\}$ of points in Y with $y_n \geqq y_m$ for $n \geqq m$ these are equivalent: (α) $\{y_n\}$ has an upper bound; (β) $\lim \sigma(y_m, y_n) = 0$ as $m, n \to \infty$; (γ) $\lim_n \sigma(y_n, y_0) = 0$ for some y_0 in Y. We also assume throughout this part that *X is connected, Λ is essentially denumerable*, and that *$\{f_\lambda \mid \lambda \varepsilon \Lambda\}$ is a family of continuous functions on X to Y with $f_\mu(x) \geqq f_\lambda(x)$ for all x in X whenever $\mu \geqq \lambda$.* For the uniform topology in Y we take that determined by σ.

For convenience, a family $\{f_\lambda\}$ of functions on X to Y will be said to be *of Class* 2 if either $C \cap C_a$ is first category or $C_a = X$; for example, if σ is right or left invariant any family of homomorphisms is of Class 2, for then C_a being non-null implies $C_a = X$.

THEOREM 7. *Let $\{f_\lambda\}$ be of Class 2 and let $L = X[x \mid \{f_\lambda(x)\}$ has an*

upper bound]. *If L contains a second category subgroup of X there exists a continuous function f_0 such that* $\lim_\lambda \sigma(f_\lambda(x), f_0(x)) = 0$ *for every x in X.*

Let G be a second category subgroup contained in L. Since G^* is a non-null open and closed subgroup and X is connected, G^* must be X and so $L^* = X$. From the equivalence of (α) and (β) in (ii) and the essential denumerability of Λ, $L = C$; hence $C^* = X$ and C is second category. Applying Theorem 2, $C \cap C_a$ is second category; this, with $\{f_\lambda\}$ being of Class 2, yields $C_a = X$. Thus $C^* \cap C_a = X$ and hence, by Lemma 1, $C = X$. This, with (ii), means that for each x there is some element $f_0(x)$ such that $\lim_\lambda \sigma(f_\lambda(x), f_0(x)) = 0$. Obviously $C(f_0) = C = X$, whence, by (2) of Lemma 2, $C_0 = X$, i. e., f_0 is continuous.

COROLLARY 7.1. *Suppose each f_λ is a homomorphism, L is second category, and σ is a right (or left) invariant metric. Then there is a continuous homomorphism f_0 on X to Y such that* $\lim_\lambda \sigma(f_\lambda(x), f_0(x)) = 0$ *for all x.*

From a remark preceding Theorem 7, $\{f_\lambda\}$ is of Class 2; from the equivalence of (α) and (γ) in (ii) it is easy to see that L is a subgroup; the conclusion now results from Theorem 7 and Corollary 6.1.

COROLLARY 7.2. *Suppose that Y is a po-group* [3], *that $\{f_\lambda\}$ is of Class 2, and that $f_\lambda(x_1 + x_2) \leqq f_\lambda(x_1) + f_\lambda(x_2)$ and $f_\lambda(x) = f_\lambda(-x)$ for arbitrary x_1, x_2, x, and λ. If L is second category the conclusion of Theorem 7 follows.*

From the assumptions on $\{f_\lambda\}$ and the fact that Y is a po-group clearly L is a subgroup; the conclusion is therefore immediate.

Now suppose that X is a real (or complex) linear topological space (hence connected), and that Y is a similar space but with a partial ordering and a pseudo-metric σ that satisfy not only (i) and (ii) above but also these: (iii) Y is a po-group, (iv) σ is right invariant, and (v) $y_1 \geqq y_2 \geqq -y_1$ implies $\sigma(y_1, \theta) \geqq \sigma(y_2, \theta)$. (Let S be a measure space with the measure of S finite; the space (M) of essentially bounded measurable real functions with the Fréchet metric and the spaces L_p, $p \geqq 1$, are examples of such Y.) Setting $\| y \| = \sigma(y, \theta)$ we note that $\| y_1 - y_2 \| = \sigma(y_1, y_2)$, $\| y \| = \| -y \|$, and $\| y_1 + y_2 \| \leqq \| y_1 \| + \| y_2 \|$.

COROLLARY 7.3. *Let $\{f_\lambda\}$ be such that* (1) $f_\lambda(x) \geqq \theta$ *for all λ and x,* (2) $f_\lambda(x_1 + x_2) \leqq f_\lambda(x_1) + f_\lambda(x_2)$, (3) $f_\lambda(x) = f_\lambda(-x)$, *and* (4)

$\| f_\lambda(x)/k \| = \| f_\lambda(x/k) \|$ for $\lambda \varepsilon \Lambda$, $x \varepsilon X$, and any positive integer k. If L is second category the conclusion of Theorem 7 follows.

From Corollary 7.2 it is sufficient to show that $\{f_\lambda\}$ is of Class 2. To do this we first observe that if $C_a \ni \theta$ then $C_a = X$. For clearly $f_\lambda(x) - f_\lambda(a) \leqq f_\lambda(x-a)$; and $f_\lambda(a) \leqq f_\lambda(a-x) + f_\lambda(x) = f_\lambda(x-a) + f_\lambda(x)$ whence $-f_\lambda(x-a) \leqq f_\lambda(x) - f_\lambda(a)$. From (v) we then have $\| f_\lambda(x) - f_\lambda(a) \| \leqq \| f_\lambda(x-a) \|$, and hence $C_a = X$ if $C_a \ni \theta$. To complete the proof it will be shown that $C_a \ni \theta$ when $C^* \cap C_a$ is non-null.

Let $\epsilon > 0$ be given and choose $\delta > 0$ so that $\| \alpha y \| < \epsilon/2$ whenever $\| y \| < \delta$ and α is a scalar with $|\alpha| < \delta$. Choose $a \varepsilon C^* \cap C_a$. Since $a \varepsilon C_a$, there is a nucleus U in X such that $\| f_\lambda(x+a) - f_\lambda(a) \| < \delta$ for x in U and λ sufficiently large. Thus

$$(*) \qquad \| \alpha f_\lambda(x+a) - \alpha f_\lambda(a) \| < \epsilon/2 \text{ for } x \varepsilon U,\ |\alpha| < \delta,\ \lambda \text{ large.}$$

Now choose $b \varepsilon C \cap (U+a)$. Since $b \varepsilon C$ implies $\| f_\lambda(b) - f_{\lambda_1}(b) \| < \delta$ for all $\lambda \geqq$ some λ_1, so that $\| \alpha f_\lambda(b) - \alpha f_{\lambda_1}(b) \| < \epsilon/2$ for $|\alpha| < \delta$ and $\lambda \geqq \lambda_1$, and since $\| \alpha f_{\lambda_1}(b) \| \to 0$ as $|\alpha| \to 0$, we can choose a positive integer k such that $1/k < \delta$ and $\| f_\lambda(b)/k \| < \epsilon$ for large λ. On the other hand, $b \varepsilon U + a$, $1/k < \delta$, and (*) give us $\| f_\lambda(b)/k - f_\lambda(a)/k \| < \epsilon/2$. Combining the last two inequalities furnishes $\| f_\lambda(a)/k \| < 2\epsilon$ for large λ. Using (*) again, we also have $\| f_\lambda(x+a)/k - f_\lambda(a)/k \| < \epsilon/2$ for x in U and λ large. At this point we note that $\| f_\lambda(x/k) \| \leqq \| f_\lambda(x+a)/k - f_\lambda(a)/k \| + 2 \| f_\lambda(a)/k \|$ holds for any x, a, k, and λ, as the reader can easily verify from hypotheses (3) and (4) and the properties of $\| y \|$. Hence $\| f_\lambda(x/k) \| < 5\epsilon$ is true for x in U and λ large. Since kx is continuous in x there is a nucleus U' such that $kU' \subset U$. For large λ and x' in U' we now have $\| f_\lambda(x') \| < 5\epsilon$, so that $\theta \varepsilon C_a$, ending the proof.

Corollary 7.3 can be considered as a slight extension of Yosida's abstraction [19, Theorem 1] of the Banach theorem on convergence almost everywhere as formulated and proved by Mazur and Orlicz [11]. In using Yosida's theorem or Corollary 7.3 to obtain any concrete individual ergodic theorem, as some authors have done, the essential difficulty is that of establishing that L is second category; this is done, for example, in the paper "On the ergodic theorem," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 538-549, by Nelson Dunford and D. S. Miller in their Lemma 6.

Remarks on completely additive set functions. Let X be a σ-field of sets in some space S and $m(x)$ a finite non-negative completely additive real

function on X; with the pseudo-metric $\rho(x_1, x_2) = m((x_1 - x_2) \cup (x_2 - x_1))$, X is a complete pseudo-metric space. Let G be an additive abelian group having an invariant pseudo-metric and let $F(m)$ be the space of functions defined and additive on X to G and absolutely continuous with respect to m (i.e., continuous with respect to ρ at the null set). Each f in $F(m)$ is completely additive. Moreover, as is easily verified, the set C_e for any subset H of $F(m)$ is either vacuous or all of X; in particular, taking H to consist of an arbitrary single element f, C_e for f is the set of continuity points of f, which shows that each f in $F(m)$ is continuous with respect to ρ. Let $\{f_n\}$ be any countable sequence in $F(m)$, so that $C_a = C_e$. If C for $\{f_n\}$ is second category, Theorem 3 establishes that C_a is second category; hence $X = C_a = C_e$, which is the Lebesgue-Hahn-Saks theorem in the form given by Saks [17]. In particular, if $C = X$ then $C_e = X$; if moreover G is complete in its pseudo-metric then $\lim_n f_n(x)$ exists on X to G and by (3) of Theorem 3 must be in $F(m)$, establishing another result given by Saks [17]. (Abstractions of Saks' theorems have been proved in this manner by Alexiewicz [20].)

Given the σ-field X and any denumerable sequence $\{f_n\}$ of completely additive finite real (or complex) functions on X, let ν_n be the total variation function of f_n, let $k_n = 1/2^n(1 + \nu_n(S))$, and set $\bar{m}(x) = \sum_{i}^{\infty} k_n \nu_n(x)$; it is easily seen that $\bar{m}$ is an m of the type above and that $\{f_n\} \subset F(\bar{m})$. If the set C for $\{f_n\}$ is X then $C_e = C_a = X$ by what was just proved; if we set $f_0(x) = \lim_n f_n(x)$ we then have, by (2) of Lemma 2, that $C_0 \supset C(f_0)^0 \cap C_e = X$, and hence f_0 is continuous on X. Being obviously additive f_0 is then in $F(\bar{m})$ and thus is completely additive. This result is due to Nikodym [14] and the present proof to Saks [17]. We may also note in this case that since $C_a = C_e = X$ the null set is an element of C_e and hence $f_n(x) \to 0$ uniformly in n as $\bar{m}(x) \to 0$; if $\{x_k\}$ is any denumerable disjoint sequence in X it is now clear that $\lim \sum_{i=k}^{\infty} f_n(x_i) = 0$ uniformly in n as $k \to \infty$. This is a recent theorem of Doubrovsky [4].

Reverting to groups we make a final remark that a recent result of Gottschalk's [7] can be given the following form, using Theorems 3 and 4 and part of Gottschalk's proof. Let X be a group and also a locally complete pseudo-metric space. Suppose the center of X is dense in X and the group product xy is continuous in x for each y. Then xy is continuous in (x, y).

TULANE UNIVERSITY.

REFERENCES.

[1] S. Banach, "Théorème sur les ensembles de première catégorie," *Fundamenta Mathematicae*, vol. 16 (1930), pp. 395-398.

[2] ———, "Über metrische Gruppen," *Studia Mathematica*, vol. 3 (1931), pp. 101-113.

[3] G. Birkhoff, *Lattice Theory* (Rev. Ed.), American Mathematical Society Colloquium Publications, vol. 25, New York, 1948.

[4] V. Doubrovsky, "On some properties of completely additive set functions etc., *Recueil Mathématique*, New Series, vol. 20(62) (1947), pp. 317-329.

[5] N. Dunford, "On continuous mapping," *Annals of Mathematics*, vol. 41 (1940), pp. 639-661.

[6] *Encyclopädie der Mathematischen Wissenschaften*, vol. II, 3, II, Leipzig, 1923-7.

[7] W. H. Gottschalk, "Transitivity and equi-continuity," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 982-984.

[8] H. Hahn, *Reelle Funktionen*, Leipzig, 1932.

[9] F. Hausdorff, *Mengenlehre*, 3. Auflage, Berlin and Leipzig, 1935.

[10] C. Kuratowski, "Sur les fonctions représentables analytiquement et les ensembles de première catégorie," *Fundamenta Mathematicae*, vol. 5 (1924), pp. 75-86.

[11] S. Mazur and W. Orlicz, "Über Folgen linearer Operationen," *Studia Mathematica*, vol. 4 (1933), pp. 152-157.

[12] ———, "Sur les espaces métriques linéaires (I)," *Studia Mathematica*, vol. 10 (1948), pp. 184-208.

[13] D. Montgomery, "Continuity in topological groups," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 879-882.

[14] O. Nikodym, "Sur les suites convergences de fonctions parfaitement additives d'ensemble abstrait," *Monatshefte für Mathematik und Physik*, vol. 40 (1933), pp. 965-970.

[15] W. F. Osgood, "Non-uniform convergence and the integration of series term by term," *American Journal of Mathematics*, vol. 19 (1897), pp. 155-190.

[16] B. J. Pettis, "On the continuity and openness of homomorphisms in topological groups," *Annals of Mathematics*, vol. 52 (1950), pp. 293-308.

[17] S. Saks, "Addition to the note 'On some functionals'," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 965-970.

[18] A. Weil, "Sur les Espaces à Structure Uniforme et sur la Topologie Générale," *Actualités Scientifiques et Industrielles*, no. 551, Paris, 1937.

[19] K. Yosida, "An abstract treatment of the individual ergodic theorem," *Proceedings of the Imperial Academy (Tokyo)*, vol. 16 (1940), pp. 280-284.

[20] A. Alexiewicz, "On sequences of operations (I)," *Studia Mathematica*, vol. 11 (1949), pp. 1-30.

AN ITERATION FORMULA FOR FREDHOLM INTEGRAL EQUATIONS OF THE FIRST KIND.*

By L. Landweber.

1. Introduction. Neumann's Method of solving Fredholm integral equations of the *second* kind by iteration is of great practical and theoretical value. For Fredholm integral equations of the *first* kind, on the other hand, Hellinger and Toeplitz [1] remark that a method of solution by iteration is not available.

Physical problems often lead to an integral equation of the first kind to which a good first approximation may be derived by physical reasoning. An example of this is the problem of determining an axial source-sink or doublet distribution which would yield the axially-symmetric potential flow about a body of revolution in a uniform stream. This problem leads to an integral equation of the *first* kind, $\frac{1}{2} = \int_0^1 [(x-t)^2 + y(x)^2]^{-3/2} m(t)\,dt$, where the axis of the body coincides with the x-axis from $x = 0$ to $x = 1$, $y(x)$ is a known function, representing the ordinates of the intersection of the given surface with a meridian plane and $m(x)$ is an unknown function, representing the distribution of the doublet strength per unit length along the axis. A well-known, excellent, first approximation to the doublet distribution for elongated bodies of revolution is [2] $m_0(x) = [y(x)]^2/4$. In cases such as this it would be highly desirable to have a method of successive approximations for improving upon this approximation.

The theories of Schmidt and Picard furnish expressions for solutions to integral equations of the first kind. However, these expressions are of little practical value since they involve the characteristic numbers and functions of an arbitrary kernel, and the methods for obtaining these are both tedious and approximate.

It is proposed to present an iteration formula for obtaining successive approximations to the solution of Fredholm integral equations of the *first* kind, and to prove the convergence of the successive approximations under various conditions.

* Received August 10, 1950.

2. The integral equation of the first kind; theory of Schmidt and Picard. We are concerned with solutions and approximations to solutions of the integral equation of the first kind

$$f(x) = \int_a^b k(x, y) g(y) dy, \tag{1}$$

where $f(x)$ and $k(x, y)$ are given continuous real functions in $a \leqq x, y \leqq b$, and $g(y)$ is an unknown function. As is well known, (1) may be transformed into the integral equation with a symmetric kernel,

$$F(x) = \int_a^b K(x, y) g(y) dy, \tag{2}$$

where

$$K(x, y) = \int_a^b k(t, x) k(t, y) dt \tag{3}$$

and

$$F(x) = \int_a^b k(y, x) f(y) dy. \tag{4}$$

A theory due to E. Schmidt [3] shows that there exists a set $\{\lambda_i\}$ of positive characteristic numbers, which may be supposed arranged in increasing order of magnitude, and corresponding adjoint sets $\phi_i(x)$ and $\psi_i(x)$ of real, continuous, orthonormalized characteristic functions, $(i = 1, 2, \cdots)$, such that

$$\phi_i(x) = \lambda_i \int_a^b k(x, y) \psi_i(y) dy, \qquad \psi_i(x) = \lambda_i \int_a^b k(y, x) \phi_i(y) dy. \tag{5}$$

It will be convenient, hereafter, to employ the customary operator notation for integral transforms, viz., $kg \equiv \int_a^b k(x, y) g(y) dy$, $Kg \equiv \int_a^b K(x, y) g(y) dy$; furthermore, since the range of variation and the integration limits will always be from a to b, specific reference to these limits will be omitted and we will frequently write integrals in an abbreviated form, viz., $\int_a^b f(x)\phi_i(x) dx \equiv \int f\phi_i$.

If the kernel $k(x, y)$ is degenerate, the number of characteristic functions is finite and they can be found by a well known procedure [4]. If $f(x)$ is expressible in the form $f(x) = \sum_{i=1}^n a_i \phi_i(x)$, the solution of (1) is

$$g(x) = \sum_{i=1}^n \lambda_i a_i \psi_i(x), \qquad a_i = \int f\phi_i. \tag{6}$$

If $f(x)$ is not of the above form, then (6) gives the best approximate solution of (1) in the least square sense, as can easily be shown. If the kernel $k(x, y)$

is non-degenerate, the sets $\{\lambda_i\}$, $\{\phi_i(x)\}$ and $\{\psi_i(x)\}$ are infinite. Since the degenerate case is readily disposed of, only the non-degenerate case will be considered hereafter.

These characteristic numbers and adjoint functions have several properties which will be required in the following:

a) λ_i^2 and $\psi_i(x)$ are characteristic numbers and functions of $K(x, y)$ [5], i. e.,

$$\psi_i = \lambda_i^2 K \psi_i. \tag{7}$$

b) A positive lower bound for the set $\{\lambda_i\}$ is given by the inequality [3]

$$1/\lambda_1^2 < \int\int k^2(x, y)\, dx dy. \tag{8}$$

c) Expansion Theorems. Every function $f(x)$ of the form (1), where $g(y)$ is any piecewise-continuous function, can be expanded in the absolutely and uniformly convergent series [5]

$$f(x) = \sum_{i=1}^{\infty} a_i \phi_i(x)\,; \qquad a_i = \int f\phi_i = (\int g\psi_i)/\lambda_i. \tag{9}$$

Every function $F(x)$ of the form (4), where $f(x)$ is any piecewise-continuous function, can be expanded in the absolutely and uniformly convergent series

$$F(x) = \sum_{i=1}^{\infty} c_i \psi_i(x)\,; \qquad c_i = \int F\psi_i = (\int f\phi_i)/\lambda_i. \tag{10}$$

If f is the same function in (9) and (10), the relations between the "Fourier" coefficients may be written

$$c_i = \int F\psi_i = (\int f\phi_i)/\lambda_i = (\int g\psi_i)/\lambda_i^2. \tag{11}$$

In general a solution of (1) does not exist. A theorem due to E. Picard [6] states that, *if the orthogonal set ϕ_i is complete, a solution of the integral equation* (1) *exists if and only if the series*

$$\sum_{i=1}^{\infty} \lambda_i^2 a_i^2, \; a_i = \int f\phi_i \tag{12}$$

is convergent.

In the Schmidt-Picard theory, the solution of (1) is intimately related to the sequence

$$\bar{g}_n = \sum_{i=1}^{n} \lambda_i a_i \psi_i(x), \qquad n = 1, 2, \cdots \tag{13}$$

as is expressed in the following theorems:

THEOREM 1. *The sequence* $\{k\bar{g}_n\}$ *converges in the mean to* $f(x)$ *if and only if the set* $\{\phi_i\}$ *is complete relative to* $f(x)$. *The sequence converges uniformly to* $f(x)$, *if a piecewise-continuous solution of the integral equation* (1) *exists.*

THEOREM 2. *If a piecewise-continuous solution* $g(x)$ *of* (1) *exists, the sequence* $\{\bar{g}_n\}$ *converges in the mean to* $g(x)$ *if and only if the set* $\{\psi_i\}$ *is complete relative to* $g(x)$. *If* $g(x)$ *is of the form* $\int k(y, x)h(y)dy$, *where* $h(y)$ *is any piecewise-continuous function, then the sequence* $\bar{g}_n$ *converges uniformly to* $g(x)$.

The completeness conditions on the sequences $\{\phi_i\}$ and $\{\psi_i\}$ in Theorems 1 and 2 refer to the so-called completeness relations

$$(14)\qquad \int f^2 = \sum_{i=1}^{\infty} a_i^2, \quad a_i = \int f\phi_i \quad \text{and} \quad \int g^2 = \sum_{i=1}^{\infty} b_i^2, \quad b_i = \int g\psi_i.$$

The phrase "complete relative to $f(x)$" in Theorem 1 signifies that (14) need be satisfied only by the particular function $f(x)$, a condition which is considerably weaker than the assumption that the set $\{\phi_i\}$ is complete relative to a class of functions. Similarly (14) is assumed to apply only to the particular function $g(x)$ in Theorem 2.

The first part of Theorem 1 is of especial interest since it indicates that with increasing n, the error due to the assumption of $\bar{g}_n(x)$ as an approximate solution of (1) diminishes in a least square sense, even if a solution of (1) does not exist. However the disagreeable possibility exists that, beyond some value of n, the error may accumulate and increase at some values of x. Nevertheless, even in this case, such a sequence may give useful successive approximations in a particular problem, if the errors are observed at each step, and the approximations stopped when the error exceeds an acceptable value at any point.

The second part of Theorem 1 asserts that, for sufficiently large n, $\bar{g}_n$ satisfies the integral equation (1) as closely as desired. It is noteworthy that no assumptions are made with regard to the convergence of the sequence $\{\bar{g}_n\}$. Indeed, Theorem 2 shows that an additional condition is necessary to assure even convergence in the mean.

The expression (13) for $\bar{g}_n$, however, is of little practical value since it is expressed in terms of the characteristic numbers and functions of the kernel $k(x, y)$. Principally for these reasons the Fredholm integral equation of the first kind has been considered to be of little value [7]. On the other hand, another readily calculable sequence of functions $\{g_n(x)\}$ will be defined, which, it will be shown, has properties relative to a solution of the integral equation (1) identical to those of $\bar{g}_n(x)$.

3. The iteration formula. Let us now extend the operator notation, denoting $K^r g \equiv \int \cdots \int K(x, y_r) K(y_r, y_{r-1}) \cdots K(y_2 y_1) g(y_1)\, dy_r dy_{r-1} \cdots dy_1$. This notation is appropriate since the relation $K^r(K^s g) \equiv K^{r+s} g$ is satisfied, as is easily verified.

Let $g_0(x)$ be an assumed, approximate, piecewise-continuous solution of the integral equation (1). Then a set of continuous functions $g_1(x), g_2(x), \cdots$ is defined by the iteration formula

$$g_n = g_{n-1} + F - K g_{n-1} \tag{15}$$

where K and F are the functions defined in equations (3) and (4). The convergence of this sequence of functions and the applicability of its members as successive approximations to a solution of the integral equation (1) is the subject of the subsequent discussion.

The recurrence formula (15) can be readily solved for g_n in terms of g_0. First put

$$\gamma_n = g_n - g_{n-1}. \tag{16}$$

Then

$$g_n = g_0 + \sum_{i=1}^{n} \gamma_i \tag{17}$$

and also (15) may be written as

$$\gamma_n = F - K g_{n-1}. \tag{18}$$

Thus the γ_n's are not only the differences between successive g_n's but also serve as measures of the errors corresponding to the g_n's as approximate solutions of the iterated integral equation (2). Now, from (18), we have $\gamma_n - \gamma_{n-1} = - K\gamma_{n-1}$ or, in operation notation, $\gamma_n = (1 - K)\gamma_{n-1}$. Hence, since the operator K satisfies the associative laws of multiplication, we obtain

$$\gamma_n = (1 - K)^{n-1} \gamma_1, \tag{19}$$

where $(1 - K)^{n-1}$ is to be formally expanded by the binomial theorem before operating on γ_1. Substituting for the γ_1 in equation (17) from equation (19), and performing the indicated summation, we obtain

$$g_n = g_0 + \{[1 - (i - K)^n]/K\}(F - K g_0), \tag{20}$$

where, in the fractional operator, $(1 - K)^n$ is to be expanded by the binomial theorem and a factor K in the numerator cancelled with the denominator before operating on $(F - K g_0)$.

If the sequence $\{g_n(x)\}$ converges uniformly, it is clear from (15), that $\lim g_n$ is a solution of the iterated integral equation (2). However, since an

integral equation of the first kind has a solution only under special circumstances, $\{g_n(x)\}$ may not converge uniformly, and indeed may not converge at all. Nevertheless the g_n's may serve as useful approximations to a solution of (1) and (2) as will be evident on the basis of the convergence theorems in the next section.

4. Convergence theorems. It will be assumed hereafter that

$$\int_a^b \int_a^b k^2(x, y)\,dxdy \leqq 2. \tag{21}$$

This is no restriction since $k(x, y)$ can always be modified, so as to satisfy (21), by multiplying (1) by a suitable factor and, in the right member of the equation, incorporating the factor into the kernel.

The convergence theorems will first be stated and discussed briefly before their proofs are presented.

THEOREM 3. *The sequence $\{Kg_n\}$ converges uniformly to $F(x)$.*

Theorem 3 is very strong. Without any restrictive assumptions about completeness, the existence of a solution, or the convergence of the sequence $\{g_n\}$, it asserts that, for sufficiently large n, g_n satisfies the iterated integral equation (2) as closely as desired. Basically, however, our interest is in the integral equation (1), rather than with (2). Concerning the suitability of the g_n's as approximate solutions of (1) we have the weaker theorems.

THEOREM 4. *The sequence $\{kg_n\}$ converges in the mean to $f(x)$ if and only if the set $\{\phi_i\}$ is complete relative to $f(x)$. The sequence converges uniformly to $f(x)$ if a piecewise-continuous solution of the integral equation* (1) *exists.*

It will now be supposed that the 0-th approximation $g_0(x)$ is chosen of the form

$$g_0(x) = \int k(y, x) h(y)\,dy, \tag{22}$$

where $h(y)$ is any piecewise-continuous function. The special case $h(y) \equiv 0$ is also allowed. Concerning the convergence of the sequence $\{g_n\}$ we then have

THEOREM 5. *If a piecewise-continuous solution $g(x)$ of* (1) *exists, the sequence $\{g_n\}$ converges in the mean to $g(x)$ if and only if the set $\{\psi_i\}$ is complete relative to $g(x)$. If $g(x)$ is of the form $\int k(y, x) h(y)\,dy$, where $h(y)$ is any piecewise-continuous function, then the sequence $\{g_n\}$ converges uniformly to $g(x)$.*

It should be noted that Theorems 4 and 5 are identical, word for word, with Theorems 1 and 2 except for the substitution of g_n for $\bar{g}_n$. Hence the remarks concerning the suitability of the $\bar{g}_n$'s as approximations to a solution of the integral equation (1) are applicable to the g_n's as well.

In order to prove the foregoing theorems it is first convenient to establish several lemmas. Put

$$(23) \qquad F_n(x) \equiv Kg_n, \qquad f_n(x) \equiv kg_n.$$

The "Fourier" coefficients of F_n, f_n and g_n then satisfy the relations

$$(24) \qquad c_{in} = \int F_n\psi_i = (\int f_n\phi_i)/\lambda_i = (\int g_n\psi_i)/\lambda_i^2.$$

We then have

LEMMA 1. *$F_n(x)$ and $f_n(x)$ can be expanded in the absolutely and uniformly convergent series*

$$(25) \quad F_n(x) = \sum_{i=1}^{n} c_{in}\psi_i(x), \qquad f_n(x) = \sum_{i=1}^{n} \lambda_i c_{in}\phi_i(x), \qquad n = 0, 1, 2, \cdots.$$

If $g_0(x)$ is chosen of the form (22), then also $g_n(x)$ may be expanded in the absolutely and uniformly convergent series

$$(26) \qquad g_n(x) = \sum_{i=1}^{n} \lambda_i^2 c_{in}\psi_i(x), \qquad n = 0, 1, 2, \cdots.$$

Proof. It is clear, from their definitions in (23), that the expansion theorems apply to $F_n(x)$ and $f_n(x)$ and consequently the series (25) converge as stated in the lemma. In the case of the g_n's, it can be shown successively, from the iteration formula (15), that $g_1(x), g_2(x), \cdots$ are of the same form as $g_0(x)$. Thus we have

$$(27) \qquad g_1 = g_0 + F - Kg_0.$$

But $g_0 = \int k(y, x)h(y)dy$; from (4), $F = \int k(y, x)f(y)dy$; and from (3), (23), $Kg_0 = \int k(y, x)f_0(y)dy$. Hence (27) becomes $g_1 = \int k(y, x)[h(y) + f(y) - f_0(y)]dy$. Hence the expansion theorem is applicable to $g_n(x)$ and the series (26) also converge, as stated.

LEMMA 2.

$$(28) \qquad c_{in} - c_i = \mu_i^n(c_{i0} - c_i),$$

where $c_i = \int F\psi_i$, and the sequence μ_i is such that

$$(29) \qquad |\mu_i| < 1, \quad \mu_{i+1} \geqq \mu_i \text{ and } \lim_{i\to\infty} \mu_i = 1, \qquad i = 1, 2, \cdots.$$

Proof. We obtain, from (15) and (7), $\int g_n\psi_i = (1 - 1/\lambda_i^2)\int g_{n-1}\psi_i$

$+\int F\psi_i$. Put $\mu_i = 1 - 1/\lambda_i^2$. Then, by successive reduction, we obtain $\int g_n\psi_i = \mu_i^n \int g_0\psi_i + \lambda_i^2(1-\mu_i^n)\int F\psi_i$, which by (11) and (24), is seen to be equivalent to (28). Furthermore, from (8) and (21), we obtain $0 < 1/\lambda_i^2 < \int\int k^2(x,y)\,dxdy \leqq 2$, or $-1 < \mu_i < 1$. Thus, since the sequence $\{\lambda_i\}$ increases monotonically to infinity, it is seen that (29) is also satisfied. This completes the proof of Lemma 2.

LEMMA 3.

$$\lim_{n\to\infty}\sum_{i=1}^{\infty}(c_{in}-c_i)^2 = \lim_{n\to\infty}\sum_{i=1}^{\infty}\lambda_i^2(c_{in}-c_i)^2 = 0. \tag{30}$$

If a solution $g(x)$ of (1) or (2) exists, then also

$$\lim_{n\to\infty}\sum_{i=1}^{\infty}\lambda_i^4(c_{in}-c_i)^2 = 0. \tag{31}$$

Proof. We first note that the series $\sum_{i=1}^{\infty}(c_{i0}-c_i)^2$ converges since we have, from Bessel's inequality $\sum_{i=1}^{\infty}(c_{i0}-c_i)^2 \leqq \int(F_0-F)^2$; hence, $\sum_{i=1}^{\infty}(c_{in}-c_i)^2 = \sum_{i=1}^{\infty}\mu_i^{2n}(c_{i0}-c_i)^2$ is uniformly convergent in n, by (28) and the comparison test. Consequently, $\lim_{n\to\infty}\sum_{i=1}^{\infty}(c_{in}-c_i)^2 = \sum_{i=1}^{\infty}\lim_{n\to\infty}\mu_i^{2n}(c_{i0}-c_i)^2 = 0$. Similarly, applying Bessel's inequality to $f_0 - f$, and then to $g_0 - g$, when $g(x)$ is assumed to exist, we obtain (30) and (31), as desired.

LEMMA 4. *If the series $\Gamma_0(x) = \sum_{i=1}^{\infty} w_i(x)$, where the $w_i(x)$ are continuous functions, is absolutely and uniformly convergent, and if $\Gamma_n(x) = \sum_{i=1}^{\infty}\mu_i^n w_i(x)$, $n = 0, 1, 2, \cdots$, where μ_i satisfies condition (29), then the sequence $\Gamma_n(x)$ converges uniformly to 0.*

Proof. From the hypotheses on μ_i we have, for some sufficiently large r, $\mu_r \geqq |\mu_i|$, $r > i$. Also, considering the series for $\Gamma_0(x)$, given an $\epsilon > 0$, r can be chosen so large, and independent of x, that $\sum_{i=r+1}^{\infty}|w_i| < \epsilon/2$. Let r be chosen so that both conditions are satisfied. Further, we have $\sum_{i=1}^{r}|w_i| \leqq \sum_{i=1}^{\infty}|w_i| < M$, where M is an upper bound independent of x. Choose N sufficiently large so that $\mu_r^n < \epsilon/(2M)$ for $n > N$. Then $|\Gamma_n| \leqq \sum_{i=1}^{r}|\mu_i^n w_i| + \sum_{i=r+1}^{\infty}|\mu_i^n w_i| \leqq \mu_r^n M + \epsilon/2 < \epsilon$, when $n > N(\epsilon)$, as we wished to prove.

LEMMA 5. *If $G_n(x)$ can be expanded in a uniformly convergent series*

$$G_n(x) = \sum_{i=1}^{\infty} e_{in}\theta_i(x), \qquad n = 0, 1, 2, \cdots \tag{32}$$

in terms of the real, continuous, orthonormalized functions $\theta_i(x)$, $i = 1, 2, \cdots$ and if $G(x)$ is piecewise-continuous, with $e_i = \int G\theta_i$, then necessary and sufficient conditions for the sequence $G_n(x)$ to converge in the mean to $G(x)$ are that $\int G^2 dx = \sum_{i=1}^{\infty} e_i^2$ and $\lim_{n\to\infty} \sum_{i=1}^{\infty} (e_{in} - e_i)^2 = 0$.

Proof. Since the series (32) is uniformly convergent, we have $\int GG_n = \sum_{i=1}^{\infty} e_{in} \int G\theta_i = \sum_{i=1}^{\infty} e_{in}e_i$, and similarly $\int G_n{}^2 = \sum_{i=1}^{\infty} e_{in}{}^2$. Hence

$$\int (G_n - G)^2 = \int G^2 + \sum_{i=1}^{\infty} (e_{in} - e_i)^2 - \sum_{i=1}^{\infty} e_i^2. \tag{33}$$

Now suppose the conditions of the lemma to be satisfied. Then $\int (G_n - G)^2 = \sum_{i=1}^{\infty} (e_{in} - e_i)^2$ and consequently by hypothesis, $\lim \int (G_n - G)^2 = 0$. This proves the first part of the lemma.

Now suppose that $\lim \int (G_n - G)^2 = 0$. From (33), $\int G^2 dx \leqq \sum_{i=1}^{\infty} e_i^2 + \int (G_n - G)^2$ for all n. Hence $\int G^2 \leqq \sum_{i=1}^{\infty} e_i^2$. But, by Bessel's inequality, $\int G^2 \geqq \sum_{i=1}^{\infty} e_i^2$. Hence $\int G^2 = \sum_{i=1}^{\infty} e_i^2$. Then, from (33), $\sum_{i=1}^{\infty} (e_{in} - e_i)^2 = \int (G_n - G)^2$, whence $\lim \sum_{i=1}^{\infty} (e_{in} - e_i)^2 = 0$. This completes the proof.

We can now proceed to the proof of the convergence theorems.

Proof of Theorem 3. By the expansion theorem and (11) and (24), the series $F_n - F = \sum_{i=1}^{\infty} (c_{in} - c_i)\psi_i$, $n = 0, 1, 2, \cdots$ are absolutely and uniformly convergent in x. Hence, by Lemma 2, the series $\sum_{i=1}^{\infty} \mu_i{}^n (c_{i0} - c_i)\psi_i$ are also absolutely and uniformly convergent in x. Hence the conditions of Lemma 4 are satisfied and the sequence $\{F_n - F\}$ converges uniformly to zero; or by (23), $\{Kg_n\}$ converges uniformly to F, as we wished to prove.

Proof of Theorem 4. By Lemmas 1 and 3 all the conditions of Lemma 5 are satisfied by the functions $f_n(x)$ and $f(x)$. Hence by (23) the first part of the theorem, concerning the convergence in the mean of $\{kg_n\}$ to $f(x)$, is proved.

In the second part of the theorem, since $g(x)$ exists by hypothesis, the expansion theorem may be applied to $f(x)$ as well as to $f_n(x)$. Hence by (11) and Lemmas 1 and 2, the series $f_n - f = \sum_{i=1}^{\infty} \mu_i{}^n \lambda_i (c_{i0} - c_i) \phi_i(x)$, $n = 0, 1, 2, \cdots$ are absolutely and uniformly convergent in x, and the conditions of Lemma 4 are satisfied. Hence the sequence $\{f_n - f\}$ converges uniformly to zero, or, by (23), $\{kg_n\}$ converges uniformly to $f(x)$. This completes the proof.

Proof of Theorem 5. Since $g_0(x)$ is of the form (22), Lemmas 1 and 3 indicate that the conditions of Lemma 5 are satisfied by the functions $g_n(x)$ and $g(x)$. Hence the first part of the theorem, concerning convergence in the mean of $\{g_n\}$ to $g(x)$, is proved.

In the second part of the theorem, the expansion theorem is applicable to $g(x)$, by hypothesis. Hence, by (11) and Lemmas 1 and 2, the series $g_n - g = \sum_{i=1}^{\infty} \mu_i{}^n \lambda_i{}^2 (c_{i0} - c_i) \psi_i(x)$, $n = 0, 1, 2, \cdots$ are absolutely and uniformly convergent in x, and the conditions of Lemma 4 are satisfied. Hence the sequence $\{g_n\}$ converges uniformly to $g(x)$, as we wished to prove.

DAVID TAYLOR MODEL BASIN,
WASHINGTON, D. C.

REFERENCES.

[1] E. Hellinger and O. Toeplitz, "Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten," *Encyklopädie der Mathematischen Wissenschaften* II C 13 (1928), pp. 1349.

[2] M. M. Munk, "Fluid Mechanics," Part II of *Aerodynamic Theory*, edited by W. F. Durand, vol. I, Julius Springer, Berlin (1935), Division C, pp. 269, equation (6.1).

[3] E. Schmidt, "Zur Theorie der linearen und nichtlinearen Integralgleichungen," *Mathematische Annalen*, vol. 63 (1907), pp. 433-476.

[4] R. Courant und D. Hilbert, "Methoden der Mathematischen Physik," vol. 1, Kap. III, § 2.

[5] Reference [4], § 10.

[6] E. Picard, "Sur un théorème général relatif aux équations intégrales de première espèce et sur quelques problèmes de physique mathématique," *Rend. Circ. mat. Palermo*, vol. 29 (1910), pp. 79-97.

[7] A. Weinstein, "On axially symmetric flows," *Quarterly of Applied Mathematics*, vol. 5, No. 4 (1948), p. 433.

ON THE ASYMPTOTIC EVALUATION OF A CLASS OF MULTIPLE INTEGRALS INVOLVING A PARAMETER.*

By L. C. Hsu.

1. Introduction. The object of this paper is to determine the asymptotic behaviour of a class of multiple integrals of the form

$$I(\lambda) = \int_R \cdots \int \exp f(x_1, \cdots, x_n; \lambda)\, dx_1 \cdots dx_n,$$

as $\lambda \to \infty$. Here f is a real valued function of $x = (x_1, \cdots, x_n)$ defined in R and satisfying certain general conditions, R being a simply connected, finitely bounded, n-dimensional domain in the euclidean n-space.

We assume that, for every large λ, the function $f(x_1, \cdots, x_n; \lambda) = f(x; \lambda)$ attains an absolute maximum at an interior point $x = x(\lambda) = (x_1(\lambda), \cdots, x_n(\lambda))$ of R. The case in which $f = \lambda g(x_1, \cdots, x_n)$ and the function g takes a maximum at a boundary point of the closed domain R has been discussed in a previous paper [5]. But the method of [5] cannot be applied unaltered to the present case, because $x(\lambda)$ is now a variable point depending on the parameter λ. Thus in the present investigation we shall modify and extend Laplace-Haviland's method ([1], [2], [3]).

2. Notation and abbreviation. Throughout x, u, ξ, etc. denote respectively the points $(x_1, \cdots, x_n)$, $(u_1, \cdots, u_n)$, $(\xi_1, \cdots, \xi_n)$, etc. in the euclidean n-space, f_k denotes the partial derivative of $f(x; \lambda)$ with respect to x_k, and f_{ik} the second order partial derivative of f with respect to x_k and x_i. Moreover $H_k[-f]$ denotes the Hessian of $-f(x; \lambda)$ with regard to the first k variables $x_1, \cdots, x_k$, viz., $H_k[-f(x; \lambda)] = \det \| -f_{\alpha\beta}(x; \lambda) \|$, where $1 \leqq \alpha, \beta \leqq k$ $(k = 1, \cdots, n)$.

By $(x; \lambda) \to (\xi; \infty)$ we mean that x, λ tend to ξ, ∞ respectively and independently. Thus, for instance, $f(x; \lambda) = o(g(x; \lambda))$ as $(x; \lambda) \to (\xi; \infty)$ means that for any given $\epsilon > 0$, there exist a small number $\delta = \delta(\epsilon) > 0$ and a large number $N = N_{(\epsilon)} > 0$ such that $|f/g| < \epsilon$ whenever $|x_k - \xi_k| < \delta$ $(k = 1, \cdots, n)$ and $\lambda > N$. In an analogous manner we may define $f = O(g)$, $f \sim g$ as $(x; \lambda) \to (\xi; \infty)$.

As a convention we assume that any statement involving index i or k

* Received June 8, 1950.

is true for all $i, k = 1, \cdots, n$. Thus, for example, $f_k = 0$ represents not a single equation but a system of simultaneous equations $f_k(x;\lambda) = 0$ $(k = 1, \cdots, n)$.

3. Statement of theorems. For convenience we shall say that $f(x;\lambda)$ belongs to the class C^2 with parameter λ if f, f_k, f_{ik} are continuous functions with respect to the variables x and λ.

THEOREM 1. *Let $f(x;\lambda) \varepsilon C^2$ for $x \varepsilon R$ with parameter λ such that*

(i) *for every large λ the function $f(x;\lambda)$ attains an absolute maximum at an interior point $x(\lambda) = (x_1(\lambda), \cdots, x_n(\lambda))$ of R, so that $f_k = 0$, $H_k[-f] > 0$ at $x = x(\lambda)$, where $x(\lambda)$ tends to an interior point ξ of R as $\lambda \to \infty$;*

(ii) *for λ large $H_k[-f] > 0$ hold throughout R;*

(iii) *for $(x;\lambda) \to (\xi;\infty)$ we have $f_{ik}(x;\lambda) \sim f_{ik}(\xi;\lambda)$, $H_k[-f] \to \infty$ and $\log H_n[-f] = o(H_k[-f]/H_{k-1}[-f])$, where $H_0[-f] = 1$.*

Then for $\lambda \to \infty$ we have

$$\int_R \cdots \int \exp f(x;\lambda)\, dx_1 \cdots dx_n \sim \exp[f(x_1(\lambda), \cdots, x_n(\lambda);\lambda)]((2\pi)^n/H_n[-f(\xi;\lambda)])^{1/2}. \tag{1}$$

Note that the right-hand side of (1) does not involve R. As a matter of fact, our proof of Theorem 1 depends essentially on an asymptotic investigation of the integral in a small neighbourhood of $x = x(\lambda)$ or of $x = \xi$. Thus obviously the domain of integration R may be replaced by a larger domain D, provided that

(iv) *for λ large we have* l. u. b. $f(y_1, \cdots, y_n;\lambda) \leqq$ g. l. b. $f(x_1, \cdots, x_n;\lambda)$ *in which $x \varepsilon R$, $y = (y_1, \cdots, y_n) \varepsilon D - R$, where R may be an arbitrary small domain containing ξ but fixed.*

Moreover by an application of the mean value theorem for the integral calculus we easily establish the following

THEOREM 2. *If $\phi(x)$ is absolutely integrable over a finite domain D, if $f(x;\lambda)$ satisfies all the hypotheses of Theorem* 1 *together with* (iv) *in which R is a fixed neighbourhood of ξ, and if $\phi(x)$ is continuous at $x = \xi$ with $\phi(\xi) \neq 0$, then we have*

$$\int_D \cdots \int \phi(x) \exp f(x;\lambda)\, dx_1 \cdots dx_n \sim \exp[f(x(\lambda);\lambda)]\phi(\xi)((2\pi)^n/H_n[-f(\xi;\lambda)])^{1/2}. \tag{2}$$

Evidently when $n=1$ condition (iii) reduces to the following

$$\text{(iii)}^1 \qquad f''_{xx}(x;\lambda) \sim f''_{xx}(\xi;\lambda), \qquad -f''_{xx}(x;\lambda) \to \infty \text{ as } (x;\lambda) \to (\xi;\infty).$$

This condition is automatically satisfied for the simplest case $f(x;\lambda) = \lambda h(x)$ with $h'(\xi) = 0$, $h''(\xi) < 0$. Thus it is clear that the 1-dimensional case of Theorem 2 is actually a direct generalization of the classical Laplace theorem for an asymptotic integral (see [1; 277-280], [3; 78]). Similarly it is also easily seen that Lemma 1 of [5; 624] is implied by our Theorem 1 by taking $f(x_1,\cdots,x_n;\lambda) = \lambda g(x_1,\cdots,x_n)$.

4. Proof of theorems. The following Lemma 1 will play an important rôle in our proof of Theorem 1:

LEMMA 1. *Let $A = [a_{ik}]$ be a symmetrical matrix of the definite form $xAx' = \Sigma\Sigma a_{ik}x_ix_k$ in n variables. Then*

$$(3) \qquad xAx' = \sum_{k=1}^{n} (A_{k-1}^{(k-1)} \cdot A_k^{(k)})^{-1} \, (A_k^{(k)}x_k + A_{k+1}^{(k)}x_{k+1} + \cdots + A_n^{(k)} \cdot x_n)^2,$$

where x' denotes the transposed of x, and $A_0^{(0)} = 1$, $A_s^{(1)} = a_{s1}$,

$$(4) \qquad A_s^{(k)} = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \cdots & \cdots & \cdots \\ a_{k-1\,1} & \cdots & a_{k-1\,k} \\ a_{s1} & \cdots & a_{sk} \end{vmatrix}, \qquad (1 \leq k \leq s \leq n).$$

The explicit formula (3) is actually a direct consequence of Darboux's reduction of quadratic forms [4; 191-192]. It may also be proved by repeated application of Lagrange's reduction and by induction. The detailed proof will be omitted here.

In what follows we shall say that a hypersurface defined by $z = f(x_1,\cdots,x_n)$ for $x \,\varepsilon\, R$ is concave downward in R, if the second order directional derivative of f is negative definite throughout R.

LEMMA 2. *Let $f(x_1,\cdots,x_n;\lambda)$ satisfy hypotheses* (i) *and* (ii) *of Theorem* 1. *Then for every fixed large λ the hypersurface represented by $z = f(x;\lambda)$ is concave downward in R and attains its highest point at $x = x(\lambda)$.*

Let us take an arbitrary line of the euclidean n-space, say $\Gamma_{(\theta_1,\ldots,\theta_n)}$, passing through a point $(x_1^0,\cdots,x_n^0)$ of R and having directional cosines

$\cos\theta_1, \cdots, \cos\theta_n$, where θ_k is the angle made by Γ and the x_k-axis, so that $\Sigma \cos^2\theta_k = 1$ and the equation of Γ may be written as

$$(x_1 - x_1^0)/\cos\theta_1 = (x_2 - x_2^0)/\cos\theta_2 = \cdots = (x_n - x_n^0)/\cos\theta_n.$$

It is immediately found that the first and second directional derivatives of $z = f(x_1, \cdots, x_n; \lambda)$ with respect to the direction Γ_θ are as follows

$$\partial f/\partial\Gamma_\theta = \sum_{k=1}^{n} f_k \cos\theta_k, \qquad \partial^2 f/\partial\Gamma_\theta^2 = \sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}\cos\theta_i\cos\theta_k. \tag{5}$$

Obviously we may express $\partial^2 f/\partial\Gamma_\theta^2 = yHy'$, where $H = [f_{ik}]$, $y = (y_1, \cdots, y_n) = (\cos\theta_1, \cdots, \cos\theta_n)$, and $y' = \{y_1, \cdots, y_n\}$ is the transposed of y. Hence applying Lemma 1 with $a_{ik} = f_{ik}$ and using hypothesis (ii) we get

$$\partial^2 f/\partial\Gamma_\theta^2 = yHy' < 0, \tag{6}$$

for every large λ, say $\lambda > M$. Our lemma thus follows from (i) and (6).

Proof of Theorem 1. Let us now keep $\lambda (> M)$ fixed and denote

$$\int_R \cdots \int \exp f(x;\lambda)\,dR$$
$$= \int_U \cdots \int \exp f(x;\lambda)dx + \int_{R-U} \cdots \int \exp f(x;\lambda)dx = I_1(\lambda) + I_2(\lambda),$$

where U is a neighbourhood of $x(\lambda) = (x_1(\lambda), \cdots, x_n(\lambda))$ whose volume depends on λ, $R - U$ denotes the complementary region of U with respect to the whole domain R. For convenience, we take $U = U_{(\epsilon)}$ to be a hypercube having the variable length $2\epsilon(\lambda)$ on each side and containing $x(\lambda)$ as its centre. By Taylor's formula with remainder we have

(7)

$$J(\lambda) = I_1(\lambda)\exp[-f(x_1(\lambda), \cdots, x_n(\lambda); \lambda)]$$
$$= \int_U \cdots \int \exp[\tfrac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}(X_1, \cdots, X_n; \lambda)(x_i - x_i(\lambda))(x_k - x_k(\lambda))]$$
$$\times\, dx_1 \cdots dx_n,$$

where

$$|X_k - x_k(\lambda)| < |x_k - x_k(\lambda)| \leq \epsilon(\lambda). \tag{8}$$

Since all X_k depend on x we may write $X_k = X_k(x)$. Moreover for fixed x we denote

$$x_k - x_k(\lambda) = u_k, \qquad -f_{ik}(X_1, \cdots, X_n; \lambda) = b_{ik}, \qquad [b_{ik}] = B. \tag{9}$$

Then the integral $J(\lambda)$ may be written as

(10) $$J(\lambda)=\int_{-\epsilon(\lambda)}^{\epsilon(\lambda)}\cdots\int_{-\epsilon(\lambda)}^{\epsilon(\lambda)}\exp(-\tfrac{1}{2}uBu')\,du_1\cdots du_n,$$

where $u=(u_1,\cdots,u_n)$ and $u'=\{u_1,\cdots,u_n\}$.

For asymptotic evaluation of $J(\lambda)$ we need the following

LEMMA 3. *Let* $-f_{ik}(x(\lambda);\lambda)=a_{ik}$ *and* $[a_{ik}]=A$. *Then*

(11) $$\lim_{u_k\to 0}(uBu')/(uAu')=1.$$

From condition (ii) it is clear that uAu' is a positive definite form in n variables for $\lambda>M$. On the other hand, we see that the functional elements of B are continuous at $x=x(\lambda)$, i. e. $b_{ik}=-f_{ik}(X_1(x),\cdots,X_n(x);\lambda)\to -f_{ik}(x_1(\lambda),\cdots,x_n(\lambda);\lambda)=a_{ik}$ as $x\to x(\lambda)$ or correspondingly $u\to(0,\cdots,0)$. Thus when u tends to $(0,\cdots,0)$ through any path, say $u_k=\rho_k t$ as $t\to 0$, ρ_k being constants not all zero, we always have $\lim(uBu'/uAu')=\lim(\rho B\rho'/\rho A\rho')=1$ where $\rho=(\rho_1,\cdots,\rho_n)$. Our lemma is therefore proved.

Now suppose that we choose ϵ in such a way that $\epsilon(\lambda)\to 0$ as $\lambda\to\infty$. Then from (8) and (9) it follows that $u=(u_1,\cdots,u_n)\to(0,\cdots,0)$ as $\lambda\to\infty$, so that by Lemma 3 we may write

(12) $$uBu'=uAu'(1+\delta(u)),$$

where $\delta(u)\to 0$ as $\lambda\to\infty$. It is known that uAu' can be reduced to the sum of n squares by means of a congruent linear transformation $u'=Tv'$ with $T'AT=I$ so that

(13) $$uAu'=vT'ATv'=vv'=\sum_{k=1}^{n}v_k^2,$$

where $v=(v_1,\cdots,v_n)$ and the Jacobian of the transformation is given by $|\partial(u)/\partial(v)|=|T|=|A|^{-1/2}$.

Under the transformation T the neighbourhood U of u is transformed to a neighbourhood V of v. Set

(14) $$M(\epsilon)=\max_{|u_k|\leq\epsilon}\{\delta(u)\},\qquad m(\epsilon)=\min_{|u_k|\leq\epsilon}\{\delta(u)\}.$$

Clearly from (10), (12), (13) and (14) we have

$$J(\lambda)\leq|A|^{-1/2}\int_V\cdots\int\exp[(-\tfrac{1}{2}vv')(1+m(\epsilon))]\,dv_1\cdots dv_n=J(\lambda),$$

(15)

$$J(\lambda)\geq|A|^{-1/2}\int_V\cdots\int\exp[(-\tfrac{1}{2}vv')(1+M(\epsilon))]\,dv_1\cdots dv_n=J(\lambda),$$

where v_k^2 can be written explicitly by means of Lemma 1, viz.

$$(16) \qquad v_k^2 = (A_{k-1}^{(k-1)} \cdot A_k^{(k)})^{-1}(A_k^{(k)}u_k + \cdots + A_n^{(k)} \cdot u_n)^2.$$

What we shall show next is that there is a positive valued function $\epsilon(\lambda)$ satisfying the following conditions:

I) $\lim \epsilon(\lambda) = 0$ as $\lambda \to \infty$;

II) the integration limits (upper and lower limits) for each v_k of $\bar{J}$ or $\underline{J}$ tend to $\pm\infty$ as $\lambda \to \infty$;

III) $\lim (I_2(\lambda)/I_1(\lambda)) = 0$ as $\lambda \to \infty$.

We are now going to construct such a suitable $\epsilon(\lambda)$-function. For every fixed $\lambda > M$ and fixed $x \,\varepsilon\, R$, let us denote

$$(17) \qquad \mu(x;\lambda) = \min \left\{ \frac{H_1[-f]}{H_0[-f]}, \frac{H_2[-f]}{H_1[-f]}, \cdots, \frac{H_n[-f]}{H_{n-1}[-f]} \right\},$$

where $H_0[-f] = 1$. Then our $\epsilon(\lambda)$-function may simply be defined as follows

$$(18) \qquad \epsilon(\lambda) = K_0 \cdot \left(\frac{\log H_n[-f(x_1(\lambda), \cdots, x_n(\lambda); \lambda)]}{\mu(x_1(\lambda), \cdots, x_n(\lambda); \lambda)} \right)^{1/2},$$

where K_0 is a suitable positive number.

It is clear that the $\epsilon(\lambda)$-function just defined satisfies condition I) by our hypothesis (iii) that $\log H_n[-f] = o(H_k[-f]/H_{k-1}[-f])$ as $(x;\lambda) \to (\xi;\infty)$. For verification of II) one needs only to substitute $\pm\epsilon(\lambda)$ for u_k's on the right-hand side of (16), e.g. if $u_k = \pm\epsilon(\lambda)$, $u_{k+1} = \cdots = u_n = 0$, then by (17), (18) and hypothesis (iii) we have

$$\begin{aligned} v_k^2 &= (A_{k-1}^{(k-1)} \cdot A_k^{(k)})^{-1}(A_k^{(k)}\epsilon(\lambda))^2 \\ (19) \qquad &= \epsilon(\lambda)^2 H_k[-f(x(\lambda);\lambda)]/H_{k-1}[-f(x(\lambda);\lambda)] \\ &\geq K_1\epsilon(\lambda)^2\mu(x(\lambda);\lambda) = K_0^2 \cdot K_1 \log H_n[-f(x(\lambda);\lambda)] \to \infty, \end{aligned}$$

as $\lambda \to \infty$, where K_1 is a certain positive number (≥ 1) implied by (17). This shows that the range of $|v_k|$ tends to infinity with λ. Note that our previous linear transformation T is non-singular and leaves the convexity of the domain $U_{(\epsilon)}$ invariant. It is clear that the corresponding domain V $(= V_{\epsilon(\lambda)})$ of v will have no finite boundary or definite bounds as $\lambda \to \infty$, i.e. the integration limits for each v_k will become $\pm\infty$ as $\lambda \to \infty$. Thus by noticing $m(\epsilon) \to 0$, $M(\epsilon) \to 0$, as $\lambda \to \infty$) and making use of the complete probability integral, we easily obtain (cf. [5; 626])

$$\lim_{\lambda\to\infty} |A|^{\frac{1}{2}}\bar{J}(\lambda) = \lim_{\lambda\to\infty} |A|^{\frac{1}{2}}\underline{J}(\lambda) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}vv')dv_1 \cdots dv_n = (2\pi)^{\frac{1}{2}n}.$$

From (15), (7) and by hypothesis (iii) (i.e. $f_{ik}(x;\lambda) \sim f_{ik}(\xi;\lambda)$ as $(x;\lambda) \to (\xi;\infty)$) we may therefore infer that

$$(20)\qquad \begin{aligned} I_1(\lambda) &\sim \exp[f(x_1(\lambda),\cdots,x_n(\lambda);\lambda)]\,|\,A\,|^{-\frac{1}{2}}(2\pi)^{\frac{1}{2}n} \\ &\sim \exp[f(x(\lambda);\lambda]\{(2\pi)^n/H_n[-f(\xi;\lambda)]\}^{\frac{1}{2}}. \end{aligned}$$

Finally we have to verify condition III). Recall that the hypersurface $z=f(x;\lambda)$ (for fixed $\lambda > M$) defined on R is concave downward (by Lemma 2) and that U is a neighbourhood of $x(\lambda)$. Hence it is clear that in the complementary region $R-U$ the function $f(x;\lambda)$ can reach its absolute maximum only on the boundary of U. Denote by $\tilde{U}$ the boundary of U and by $\|R\|$ the volume of R. Let $h(\lambda)$ denote the maximum or the least upper bound of $\exp f(x;\lambda)$ as x varies over $\tilde{U}$. Obviously we have

$$(21)\qquad I_2(\lambda) \leqq \|R\| \cdot h(\lambda).$$

Since U is a hypercube of side-length $2\epsilon(\lambda)$ containing the point $x(\lambda)=(x_1(\lambda),\cdots,x_n(\lambda))$ as its centre, we see that every point of $\tilde{U}$ may be written in the form

$$(22)\qquad x^0=(x_1(\lambda)+\gamma_1\cdot\epsilon(\lambda),\cdots,x_n(\lambda)+\gamma_n\cdot\epsilon(\lambda)),$$

where $|\gamma_i| \leq 1$ $(i=1,\cdots,n)$ and $|\gamma_k|=1$ for at least one k $(1\leq k\leq n)$. Now suppose that x^0 is a point at which $\exp f(x;\lambda)$ takes the maximum $h(\lambda)$. We may write (cf. (7) and (8)).

$$(23)\qquad h(\lambda)=\exp[f(x(\lambda);\lambda)+\tfrac{1}{2}\epsilon^2\sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}(X_1,\cdots,X_n;\lambda)\gamma_i\gamma_k],$$

where $|X_k - x_k(\lambda)| < \gamma_k\cdot\epsilon(\lambda) \to 0$ with ϵ. Hence using Lemmas 3 and 1, we have, for λ large (cf. (16)),

$$(24)\qquad \begin{aligned} -\sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}(X_1,\cdots,X_n;\lambda)\gamma_i\gamma_k &\sim \sum_{i=1}^{n}\sum_{k=1}^{n}(-)f_{ik}(x(\lambda);\lambda)\gamma_i\gamma_k \\ &= \gamma A\gamma' = \sum_{k=1}^{n}(A_{k-1}{}^{(k-1)}\cdot A_k{}^{(k)})^{-1}(A_k{}^{(k)}\gamma_k+\cdots+A_n{}^{(k)}\cdot\gamma_n)^2 > 0. \end{aligned}$$

Note that γ_k's are constants not all zero. If $\gamma_n \neq 0$, then the positive definite form $\gamma A\gamma'$ contains the term $(A_n{}^{(n)}/A_{n-1}{}^{(n-1)})\cdot\gamma_n{}^2=(H_n[-f]/H_{n-1}[-f])\cdot\gamma_n{}^2$. Similarly if $\gamma_{k+1}=\cdots=\gamma_n=0$ but $\gamma_k\neq 0$, then $\gamma A\gamma'$ contains the positive term $(H_k[-f]/H_{k-1}[-f])\cdot\gamma_k{}^2$. Hence from (24) and (17) we see that in any case we have a positive constant, say $K_2>0$, such that

$$(25) \qquad -\sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}(X_1,\cdots,X_n;\lambda)\gamma_i\gamma_k > K_2\cdot\mu(x(\lambda);\lambda),$$

when λ is large. Thus by (20), (21), (23), (25) and (18) we have

$$\begin{aligned}
(26) & \\
& 0 < I_2(\lambda)/I_1(\lambda) \\
& \qquad < K_3 \,\|\, R \,\| \cdot A \,|^{\frac{1}{2}}/(2\pi)^{n/2} \exp\left[\tfrac{1}{2}\epsilon(\lambda)^2 \sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}(X_1,\cdots,X_n;\lambda)\gamma_i\gamma_k\right] \\
& \quad < K_4 \,|\, A \,|^{\frac{1}{2}} \exp[-\tfrac{1}{2}K_2\cdot\epsilon(\lambda)^2\cdot\mu(x(\lambda);\lambda)] \\
& \quad = K_4 \,|\, A \,|^{\frac{1}{2}} \exp\{-\tfrac{1}{2}K_2\cdot K_0^2 \log H_n[-f(x(\lambda);\lambda)]\} \\
& \quad = K_4 \,|\, A \,|^{\frac{1}{2}} \{H_n[-f(x(\lambda);\lambda)]\}^{-\frac{1}{2}K_0^2\cdot K_2} \\
& \quad = K_4 \{H_n[-f(x(\lambda);\lambda)]\}^{\frac{1}{2}(1-K_0^2\cdot K_2)},
\end{aligned}$$

where K_3, K_4 are certain positive numbers. Since the positive constant K_0 of (18) can be chosen freely, we may assume that $K_0^2K_2 > 1$. Hence the expression (26) leads at once to the conclusion that $I_2(\lambda)/I_1(\lambda) \to 0$ as $\lambda \to \infty$. Our Theorem 1 is therefore completely proved by (20) and III).

Proof of Theorem 2. Evidently the domain of integration R in the formula (1) can be replaced by D, if hypothesis (iv) is added. For we need only replace $\| R \|$ by $\| D \|$ in (21). Now let W denote an arbitrary small region (neighbourhood) of $x = \xi$. Let $J_1(\lambda)$, $J_2(\lambda)$ denote n-tiple integrals of $[\phi(x) - \phi(\xi)] \exp f(x;\lambda)$ taken over the domains W and $D - W$ respectively. We may express

$$\int_D \cdots \int \phi(x) \exp f(x;\lambda)\, dx_1 \cdots dx_n = \phi(\xi) I(\lambda) + J_1(\lambda) + J_2(\lambda),$$

where $I(\lambda)$ is defined as in § 1 (save that now the domain R is replaced by D). Since $\phi(x)$ is continuous at $x = \xi$ the difference $|\,\phi(x) - \phi(\xi)\,|$ may be made as small as we like, provided that W is sufficiently small. Thus the First Mean Value Theorem for integrals shows at once that $J_1(\lambda) = o(I(\lambda))$. Denote by $h(\lambda)$ the maximum of $\exp f(x;\lambda)$ as x varies over the boundary of W. Clearly we have

$$J_2(\lambda) \leq h(\lambda) \int_{D-W} \cdots \int |\,\phi(x) - \phi(\xi)\,|\, dx_1 \cdots dx_n,$$

which now replaces (21). Thus the same calculation as that of (26) shows that $J_2(\lambda) = o(I_1(\lambda))$, $J_2(\lambda) = o(\phi(\xi) I(\lambda))$ and this completes the proof of Theorem 2.

5. Simple consequences and discussion. It is clear that in applying the asymptotic formulas (1) and (2) we need first determine the functional solution $(x_1(\lambda), \cdots, x_n(\lambda))$ of the simultaneous equations $f_k(x_1, \cdots, x_n; \lambda) = 0$ for large λ. In case we require only the approximate values of $x_k(\lambda)$'s, we may obtain them directly by solving $f_k(x_1, \cdots, x_n; \lambda) = 0$ for some fixed large λ. Generally speaking, it is not necessarily true that $(f(x(\lambda); \lambda) - f(\xi; \lambda)) \to 0$ as $\lambda \to \infty$, so that we must not replace $\exp f(x(\lambda); \lambda)$ by $\exp f(\xi; \lambda)$ in our formulas (1) and (2), unless some further condition is imposed.

For $n = 2$ our Theorems 1 and 2 reduce to the case of double integrals. In such a case the conditions (ii) and (iii) are equivalent to the following

(ii)^1 $\Delta_1 = -f_{11}(x_1, x_2; \lambda) > 0$, $\Delta_2 = f_{11}f_{22} - f_{12}^2 > 0$ *hold in* R, *where* $\Delta_1, \Delta_2 \to \infty$ *as* $(x_1, x_2) \to (\xi_1, \xi_2)$ *and* $\lambda \to \infty$;

(iii)^1 $\log \Delta_2 = o(\Delta_1)$, $\log \Delta_2 = o(\Delta_2/\Delta_1)$, $f_{ik}(x_1, x_2; \lambda) \sim f_{ik}(\xi_1, \xi_2; \lambda)$ *as* $(x_1, x_2) \to (\xi_1, \xi_2)$ *and* $\lambda \to \infty$.

And the formula (1) becomes

$$(1)^1 \quad \int_R\int \exp f(x_1, x_2; \lambda) dR \sim \frac{(2\pi) \exp f(x_1(\lambda), x_2(\lambda); \lambda)}{\{f_{11}(\xi_1, \xi_2; \lambda) f_{22}(\xi_1, \xi_2; \lambda) - f_{12}(\xi_1, \xi_2; \lambda)^2\}^{1/2}}.$$

As an application of Theorem 1, we now consider multiple integrals of the form

$$G(\lambda) = \int_R \cdots \int \exp[\lambda g(x_1, \cdots, x_n; \lambda)] dx_1 \cdots dx_n,$$

where $g(x; \lambda)$ is a bounded function for λ large and belongs to C^2 for $x \varepsilon R$.

If furthermore $g(x; \lambda)$ satisfies the following conditions:

(a) *for every large* λ *the function* $g(x; \lambda)$ *assumes an absolute maximum at an interior point* $x(\lambda) = (x_1(\lambda), \cdots, x_n(\lambda))$ *of* R, *where* $(x_1(\lambda), \cdots, x_n(\lambda))$ *tends to an interior point* ξ *as* $\lambda \to \infty$;

(b) $H_k[-g(x; \lambda)] > c > 0$ *hold in* R *for* λ *large*;

(c) $g_{ik}(x; \lambda) \sim g_{ik}(\xi; \lambda)$ *as* $(x; \lambda) \to (\xi; \infty)$,

then obviously Theorem 1 is applicable and we have

$$(1)'' \qquad G(\lambda) \sim \frac{(\exp g(x_1(\lambda), \cdots, x_n(\lambda); \lambda))^\lambda}{\{H_n[-g(\xi_1, \cdots, \xi_n; \lambda)]\}^{1/2}} \left(\frac{2\pi}{\lambda}\right)^{n/2}.$$

For it is quite clear that in our present case the conditions $H_k[-f] \to \infty$ and $\log H_n[-f] = o(H_k[-f]/H_{k-1}[-f])$ are automatically satisfied with $f = \lambda g(x; \lambda)$. Concrete examples are easily found for illustrating the use of formula (1)''.

I wish to express my hearty thanks to the referee for valuable comments and suggestions. Finally I have to mention that the analytical method adopted here is almost the same as that already developed in my paper [6], but the results presented in this paper seem much more general than those contained in [6].

KING'S COLLEGE, ABERDEEN, SCOTLAND.

REFERENCES.

[1] D. V. Widder, *The Laplace Transform* (Princeton University Press, 1946), Chap. 7, pp. 277-280.

[2] E. K. Haviland, "On an asymptotic expression for a certain integral," *Quarterly Journal of Mathematics* (Oxford), vol. 7 (1936), pp. 152-157.

[3] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis* (New York, 1945), Kap. 5, pp. 77-82.

[4] R. F. Scott and G. B. Mathews, *Theory of Determinants and Their Applications* (Cambridge University Press, 1904).

[5] L. C. Hsu, A Theorem, "On the asymptotic behavior of a multiple integral," *Duke Mathematical Journal*, vol. 15 (1948), pp. 623-632.

[6] ———, "The asymptotic behavior of a kind of multiple integrals," to appear in the *Quarterly Journal of Mathematics* (Oxford).

THE NUMBER OF L^2-SOLUTIONS OF $x'' + q(t)x = 0$.*

By Philip Hartman.[1]

1. Let $q = q(t)$ be a real-valued continuous function for $0 \leqq t < \infty$. The differential equation

$$x'' + qx = 0 \tag{1}$$

and a homogeneous boundary condition at $t = 0$ determine a self-adjoint boundary value problem if and only if (1) is of the *Grenzpunkt* type, that is, if and only if the number of linearly independent solutions of class $L^2 = L^2(0, \infty)$ is less than 2; cf. [8]. This paper is principally concerned with conditions on q which are sufficient to assure that (1) is of *Grenzpunkt* type.

Let $N(t)$ denote the number of zeros on the interval $(0, t)$ of a non-trivial solution of (1); so that, up to an additive correction -1, 0 or 1, the integer $N(t)$ is independent of the choice of the particular solution of (1) determining $N(t)$. It was shown in [4] that a sufficient condition for (1) to be of *Grenzpunkt* type is that

$$\liminf_{t \to \infty} N(t)/t^2 < \infty. \tag{2}$$

It was also shown that, if $q^+(t) = \max(q(t), 0)$, then

$$N(t) \leqq \left(t \int_0^t q^+(s)\,ds\right)^{\frac{1}{2}} + O(1), \tag{3}$$

as $t \to \infty$. Thus, (2) holds if

$$\int^t q^+(s)\,ds = O(t^3), \tag{4}$$

as $t \to \infty$, and so, in particular, if

$$q(t) \leqq Ct^2 \tag{5}$$

holds for large values of t and some constant C.

* Received October 14, 1950.

[1] John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

The sufficiency of (5) was also obtained in [7]. The sufficient condition (5) was generalized to the conditions

$$q(t) \leqq Q^2(t) \tag{6}$$

and

$$\int^{\infty} Q^{-1}(s)\,ds = \infty, \tag{7}$$

where $Q(t) > 0$ is a function subject either to the condition that

(8) $Q^{-1}(t)$ satisfies a uniform Lipschitz condition

for large t, say, const. $\leqq t < \infty$ (cf. [5,] [6]) or to the condition that

(9) $Q(t)$ is monotone for large t

(cf. [6]). [In [5], a differential equation more general than (1) is considered; the corresponding assumptions reduce to (6), (7), (8) and (9), but the hypothesis (9) is not used in the proof.] The methods in [5], [6] are adaptations of those first used in [9] for the case $Q(t) =$ const. and then in [3] for the cases where the sign of equality holds in (6).

The generalizations of (5) to (6), (7) do not imply the sufficiency of (2), or even that of (4). It would, therefore, be of interest to obtain sufficient criteria involving $N(t)$, which would include those involving (2), hence (4), and (6), (7).

2. In this direction, a first guess might be that the t, whose square occurs in (2), can be replaced by a smooth monotone $Q(t)$ satisfying (7) (at least when lim inf is replaced by lim sup). However, this is not the case as is shown by the example $q(t) = t^2 \log^{2+4\epsilon} t$, where $0 < \epsilon \leqq \frac{1}{2}$. It is easily seen that the corresponding differential equation (1) is not of *Grenzpunkt* type. For, by known asymptotic formulae (cf. [10]), every solution $x(t)$ satisfies the estimate $x(t) = O(t^{-\frac{1}{2}} \log^{-\frac{1}{2}-\epsilon} t)$ and is therefore of class L^2, since $\epsilon > 0$. But $\pi N(t) \sim \int^{t} q^{\frac{1}{2}}(s)\,ds \sim$ const. $t^2 \log^{1+2\epsilon} t$, cf. [2]; hence,

$$\limsup_{t\to\infty} N(t)/Q^2(t) < \infty, \text{ where } Q(t) = t \log^{\frac{1}{2}+\epsilon} t$$

satisfies (7) if $\epsilon \leqq \frac{1}{2}$.

3. The general criteria for (1) to be of *Grenzpunkt* type, to be proved below, are contained in the following assertion (and its proof):

(I) *The differential equation* (1) *is of Grenzpunkt type if there exists a positive continuous function* $Q(t)$ *defined for large positive* t, *of bounded variation (on finite intervals) and satisfying*

$$(10)\qquad \limsup_{t\to\infty} \{2\int^t Q^{-1}(s)\,ds - \int^t Q^{-2}(s)\,dN(s) - Q^{-2}(t) - \int^t |\,dQ^2(s)\,|\} = \infty.$$

The proof of (I) starts with a device employed in [4]. Let $x = x(t)$ and $x = y(t)$ be two linearly independent solutions of (1) normalized by the Wronskian condition $y'x - yx' = 1$ and let $\theta(t)$ be the continuous function defined by

$$(11)\qquad \theta(t) = \arc\tan y(t)/x(t), \qquad |\,\theta(0)\,| \leqq \pi.$$

Then $\theta'(t) = (x^2(t) + y^2(t))^{-1} > 0$. The differential equation (1) is of *Grenzpunkt* type if and only if at least one of the functions $x(t)$, $y(t)$ is not of class L^2, that is, if and only if

$$(12)\qquad \int^\infty dt/\theta'(t) = \infty.$$

The inequality $1 = pp^{-1} \leqq \tfrac{1}{2}p^2 + \tfrac{1}{2}p^{-2}$, when applied to $p = (\theta'(t)/Q(t))^{\frac{1}{2}}$, gives

$$(13)\qquad 1/\theta' \geqq 2/Q - \theta'/Q^2;$$

so that

$$(14)\qquad \int_\alpha^t ds/\theta'(s) \geqq 2\int_\alpha^t Q^{-1}(s)\,ds - \int_\alpha^t Q^{-2}(s)\theta'(s)\,ds,$$

if α is sufficiently large and $\alpha < t$. The last integral can be integrated by parts:

$$(15)\qquad -\int_\alpha^t Q^{-2}(s)\theta'(s)\,ds = -\theta(t)Q^{-2}(t) + \theta(\alpha)Q^{-2}(\alpha) + \int_\alpha^t \theta(s)\,dQ^{-2}(s).$$

Since the zeros of $x(t)$, $y(t)$ separate each other, it is clear that

$$(16)\qquad |\,\theta(t) - \pi N(t)\,| \leqq \pi.$$

Hence the right hand side of (15) is minorized by π times

$$(17)\qquad -N(t)Q^{-2}(t)+N(\alpha)Q^{-2}(\alpha)$$
$$+\int_{\alpha}^{t} N(s)dQ^{-2}(s)-Q^{-2}(t)-Q^{-2}(\alpha)-\int_{\alpha}^{t} | dQ^{-2}(s)|,$$

which can be contracted into

$$(18)\qquad -\int_{\alpha}^{t} Q^{-2}(s)dN(s)-Q^{-2}(t)-Q^{-2}(\alpha)-\int_{\alpha}^{t} | dQ^{-2}(s)|.$$

Consequently,

$$\pi^{-1}\int_{\alpha}^{t} ds/\theta'(s) \geqq 2\pi^{-1}\int_{\alpha}^{t} Q^{-1}(s)\,ds$$
$$-\int_{\alpha}^{t} Q^{-2}(s)dN(s)-Q^{-2}(t)-Q^{-2}(\alpha)-\int_{\alpha}^{t} | dQ^{-2}(s)|.$$

(If $Q(s)$ is replaced by $\pi Q(s)$ and if the last inequality is multiplied by π^{-2}, the factor π^{-1} is removed from the first term on the right hand side.) If α is fixed and $t\rightarrow\infty$, then (10) implies (12), and so the assertion (I) follows.

4. Some corollaries of (I), and its proof, will now be deduced. For example, to show that (2) is sufficient for (1) to be of *Grenzpunkt* type, let $\alpha=\frac{1}{2}t$ and $Q(t)=t/\epsilon$, where $\epsilon>0$ is fixed. Then (14) becomes

$$\int_{\frac{1}{2}t}^{t} ds/\theta'(s) \geqq 2\epsilon\log 2-(4\epsilon^2t^{-2})(\theta(t)-\theta(t/2)).$$

Since $\epsilon>0$ is arbitrary, it is clear from (16) that if

$$(2')\qquad \liminf_{t\rightarrow\infty}\{N(t)-N(\tfrac{1}{2}t)\}/t^2<\infty,$$

then

$$\limsup_{t\rightarrow\infty}\int_{\frac{1}{2}t}^{t} ds/\theta'(s)>0,$$

and so (12) holds. Hence (2), or even (2′), is sufficient.

In the same manner, by replacing $Q(t)$ by $1/\epsilon$ and α by $t-1$ in (14), one proves the following:

COROLLARY 1. *The equation* (1) *is of the Grenzpunkt type if*

$$\liminf_{t\to\infty} \{N(t) - N(t-1)\} < \infty;$$

for instance, if $q(t) \leqq$ const. *holds on t-intervals which tend to* ∞ *and are of length* 1.

5. It will be seen below (cf. the Lemma, § **7**) that the proof of the following assertion implies that (6), (7), (9) are sufficient for (1) to be of the *Grenzpunkt* type.

COROLLARY 2. *Let* $Q(t) > 0$ *be defined for large t and satisfy* (7) *and* (9). *Then the assumption*

$$N(t) \leqq \int^{t} Q(s)\,ds \text{ for large } t, \tag{19}$$

hence, in particular, the assumption

$$\int^{t} q^{+}(s)\,ds \leqq t^{-1}\Big(\int^{t} Q(s)\,ds\Big)^{2} \tag{20}$$

is a sufficient condition in order that (1) *be of Grenzpunkt type.*

As an application of Corollary 2, let $Q(t)$ be a constant multiple of $t \log t$. Then

$$N(t) = O(t^{2} \log t), \tag{2''}$$

or, more particularly,

$$\int^{t} q^{+}(s)\,ds = O(t^{3} \log^{2} t), \tag{4''}$$

assures that (1) is of *Grenzpunkt* type.

In the proof of Corollary 2, it is sufficient to consider (19), since (19) is a consequence of (20), by (3) (as the function $Q(s)$ in (19) can be a multiple of that in (20), to account for the $O(1)$ term in (3)). In order that the proof include the case when (6), (7), (9) is assumed (cf. the Lemma, § **7**, below), Corollary 2 will be proved with (19) relaxed to

$$N(t) \leqq \int^{t} Q(s)\,ds + C \log Q(t) \text{ for large } t. \tag{21}$$

(Actually, the $\log Q$ can be improved to $Q^{2-\epsilon}$). In view of Corollary 1, it can be assumed that $Q(s)$ is non-decreasing.

In order to apply (I), let the function $Q(s)$ occurring in (10) be the function $Q(s)$ in (21). If the second integral in the expression $\{\ldots\}$ in (10) is integrated by parts (cf. (17) and (18)), it is seen from $dQ^{-2}(s) \leqq 0$ that $\{\ldots\}$ is not increased if $N(t)$ is replaced by a majorant, say the function on the right-hand side of (21). Hence, $\{\ldots\}$ in (10) is not less than

$$2\int^{t} Q^{-1}(s)\,ds - \int^{t} Q^{-1}(s)\,ds - C\int^{t} Q^{-3}(s)\,dQ(s)$$

$$-Q^{-2}(t) - 2\int^{t} Q^{-3}(s)\,dQ(s),$$

which is $\int^{t} Q^{-1}(s)\,ds + \frac{1}{2}CQ^{-2}(t) + \text{const.}$ Hence, (10) follows from (7).

6. The fact that (6), (7) and (8) imply that (1) is of *Grenzpunkt* type is contained in the Lemma of § 7, below, and in the following assertion:

COROLLARY 3. *Let $Q(t) > 0$ be defined for large t and satisfy* (7) *and* (8). *If, in addition, $Q(t)$ has the property that*

$$\int_{u}^{v} Q(s)\,ds \geqq 1 \tag{22}$$

whenever $t = u, v$ are successive zeros of some ($\not\equiv 0$) solution of (1), *then* (1) *is of the Grenzpunkt type.*

It is clear that (19) is a consequence of the condition involving (22), but not conversely. In the proof of Corollary 3, it can be supposed that $Q(t)$ does not tend to 0, as $t \to \infty$; for otherwise $N(t) = o(t)$, and so (1) is *Grenzpunkt* type. Similarly, it can be supposed that $\int^{\infty} Q(s)\,ds = \infty$. Let $t_1 < t_2 < \cdots$ be an unbounded sequence of t-values such that

$$\tfrac{1}{2} \leqq \int_{t_{k-1}}^{t_k} Q(s)\,ds \leqq 1 \tag{23}$$

and

$$\liminf_{k\to\infty} Q^{-2}(t_k) < \infty. \tag{24}$$

By virtue of the second inequality in (23), the hypothesis on $Q(s)$ implies

(25)
$$\int_{t_{k-1}}^{t_k} dN(s) \leqq 2.$$

The assumption (8), which is equivalent to the existence of a constant $C > 0$ satisfying

(26)
$$Q^{-2}(s) \mid dQ(s) \mid \leqq C ds$$

in the sense of Stieltjes' integration, shows that

$$\mid \log Q(v)/Q(u) \mid \leqq \int_{t_{k-1}}^{t_k} Q^{-1}(s) \mid dQ(s) \mid \leqq C \int_{t_{k-1}}^{t_k} Q(s) ds \leqq C,$$

if $t_{k-1} \leqq u < v \leqq t_k$. Hence, if $Q(u_k)$ and $Q(v_k)$ are the minimum and maximum values, respectively, of $Q(t)$ for $t_{k-1} \leqq t \leqq t_k$, then there exists a constant K satisfying

(27)
$$Q(v_k)/Q(u_k) \leqq K.$$

In order to apply (I), let the function $Q(s)$ in (10) be $1/\epsilon$ times the function $Q(s)$ occurring in (22), where $\epsilon > 0$ is a constant to be specified below, and let $t = t_k$. Then, by virtue (25), the expression $\{\ldots\}$ in (10) is minorized by

$$2\epsilon \int^{t_k} Q^{-1}(s) ds - 2\epsilon^2 \sum^{k} Q^{-2}(u_j) - \epsilon^2 Q^{-2}(t_k) - \epsilon^2 \int^{t_k} \mid dQ^{-2}(s) \mid.$$

By the first inequality in (23),

$$Q^{-2}(u_j) \leqq 2Q^{-2}(u_j) \int_{t_{j-1}}^{t_j} Q(s) ds; \text{ hence } Q^{-2}(u_j) \leqq 2K^2 \int_{t_{j-1}}^{t_j} Q^{-1}(s) ds.$$

by (27). Also, (26) implies that $\int^{t_k} \mid dQ^{-2}(s) \mid \leqq 2C \int^{t_k} Q^{-1}(s) ds$. Thus, the expression $\{\ldots\}$ in (10) is minorized by

$$\epsilon(2 - 4\epsilon K^2 - 2\epsilon C) \int^{t_k} Q^{-1}(s) ds - \epsilon^2 Q^{-2}(t_k).$$

Consequently, if $\epsilon > 0$ is sufficiently small, (10) follows from (7) and (24). This proves Corollary 3.

7. In this section, there will be proved a lemma which is of interest in itself and which shows that Corollary 2 (or 3) implies the sufficiency of

(6), (7) and (9) (or (8)). This lemma resulted from a discussion which the author had with Professor Wintner. It furnishes a simple proof of the asymptotic formula for $N(t)$ given in [2].

LEMMA. *Let $Q(t) > 0$ be a continuous function of bounded variation on the interval $0 \leqq t \leqq T$ and let n be the number of zeros of a solution $z = z(t) \not\equiv 0$ of*

$$z'' + Q^2 z = 0. \tag{28}$$

Then there exists a number τ satisfying $|\tau| \leqq \pi$ and

$$\pi n = \int_0^T Q(s)\,ds + \tau\left(\tfrac{1}{2}\int_0^T Q^{-1}(s)\,|\,dQ(s)\,| + 2\right). \tag{29}$$

In order to prove the Lemma, let $\theta = \theta(t)$ be the continuous function defined by

$$\theta(t) = \arctan\,(Qz/z'), \qquad |\,\theta(0)\,| \leqq \pi. \tag{30}$$

Since $Q > 0$ and since $\theta(t)$ increases at the t-values where $z(t) = 0$, it is clear that

$$|\,\theta(T) - \pi n\,| \leqq \pi. \tag{31}$$

From (30),

$$d\theta = (z'^2 + Q^2 z^2)^{-1}\{Q(z'^2 - zz'')\,dt + zz'\,dQ\}.$$

But $z'' = -Q^2 z$, so that

$$d\theta = Q\,dt + (z'^2 + Q^2 z^2)^{-1}(Qzz')Q^{-1}dQ.$$

Since $|\,Qzz'\,| \leqq \frac{1}{2}(z'^2 + Q^2 z^2)$, the Lemma follows from (31).

8. Although more general criteria for (1) to be of *Grenzpunkt* type result from the methods employed above, than from the method initiated in [9], the latter has the advantage of providing information about the L^2-character of the derivative $x'(t)$ of a solution $x = x(t)$ of (1) which is of class L^2. For example, it was proved in [9] that if $q(t) \leqq$ const. and $x(t)$ is of class L^2 and a solution of (1), then $x'(t)$ is of class L^2. A variant of this theorem will be obtained for the case that $q(t)$ is "nearly" a constant.

(II) *Let $q(t)$, $Q(t)$ be real-valued continuous functions for $0 \leqq t < \infty$. In addition, let $Q(t)$ satisfy* (8),

$$Q(t) \geqq \text{const.} > 0 \tag{32}$$

and

$$(33) \qquad \Big| \int_0^t q(s)\,ds \Big| \leqq Q(t).$$

Let $x = x(t)$ be of class L^2 and a solution of

$$(34) \qquad x'' + (\lambda + q)x = 0,$$

where λ is a constant. Then $x'(t)/Q(t)$ is of class L^2.

It may be mentioned that, by using a device similar to that in [6], the assumption (8) in (II) can be replaced by (9) if, in the conclusion, $x'(t)/Q(2t)$ replaces $x'(t)/Q(t)$.

The use of (II) shortens the proof of a theorem in [1] which states that if

$$(35) \qquad \limsup_{t\to\infty} \int^t |q(s)|\,ds/\log t = \lambda_0^{\frac{1}{2}} < \infty \qquad (\lambda_0 \geqq 0),$$

then (34) cannot possess a ($\not\equiv 0$) solution of class L^2 when $\lambda > \lambda_0$. For suppose that (34) possesses a solution $x = x(t) \not\equiv 0$ of class L^2, then, by (II), the function $x'/\log t$, as well as $x/\log t$, is of class L^2. On the other hand, if $\lambda > 0$, it follows that

$$|(x^2 + x'^2/\lambda)'| = |-2qxx'/\lambda| \leqq (x^2 + x'^2/\lambda)\,|q|/\lambda^{\frac{1}{2}};$$

so that

$$x^2 + x'^2/\lambda \geqq \text{const.} \exp\Big(-\int_0^t |q(s)|\,ds/\lambda^{\frac{1}{2}}\Big),$$

where const. > 0. This leads to a contradiction in view of (35) and $\lambda > \lambda_0$.

9. In order to prove (II), let (34) be multiplied by $x = x(t)$. The resulting equation can be written in the form $x'^2 = (xx')' + \lambda x^2 + qx^2$. Hence

$$(36) \qquad \int_0^t (x'/Q)^2 ds = \int_0^t (xx')'Q^{-2} ds + \lambda \int_0^t x^2 Q^{-2} ds + \int_0^t qQ^{-2}x^2 ds.$$

After an integration by parts, the first integral on the right-hand side of (36) becomes

$$xx'Q^{-2}\Big|_0^t + 2\int_0^t xx'Q^{-3} dQ.$$

The assumption (8), that is, (26), means that the last integral is

$$(37)\qquad O\Big(\int_0^t |\,xx'Q^{-1}\,|\,ds\Big) = O(1)\Big(\int_0^t x^2ds\Big)^{\frac{1}{2}}\Big(\int_0^t (x'/Q)^2ds\Big)^{\frac{1}{2}},$$

by Schwarz's inequality. Thus the assumption (32) shows that, as $t \to \infty$,

$$(38)\qquad \int_0^t (xx')'Q^{-1}ds = O(|\,xx'\,|) + O\Big(\int_0^t (x'/Q)^2ds\Big)^{\frac{1}{2}} + O(1).$$

The relation (32) also shows that

$$(39)\qquad \lambda\int_0^t x^2Q^{-2}ds = O(1),$$

as $t \to \infty$. If $I = I(t)$ denotes the integral whose absolute value occurs in (33), then the last integral in (36) is

$$Ix^2Q^{-2}\,|_0^t - 2\int_0^t Ixx'Q^{-2}ds + 2\int_0^t Ix^2Q^{-3}dQ.$$

By (26), (32) and (33), this expression is

$$O(x^2) + O\Big(\int_0^t |\,xx'Q^{-1}\,|\,ds\Big) + O\Big(\int_0^t x^2ds\Big);$$

hence, by (37),

$$(40)\qquad \int_0^t qQ^{-2}x^2ds = O(x^2) + O\Big(\int_0^t (x'/Q)\,ds\Big)^{\frac{1}{2}} + O(1).$$

Consequently, (36), (38), (39) and (40) give

$$\int_0^t (x'/Q)^2ds = O(|\,xx'\,| + x^2) + O\Big(\int_0^t (x'/Q)^2ds\Big)^{\frac{1}{2}}.$$

Thus the assertion (II) follows if it is ascertained that

$$(41)\qquad \liminf_{t\to\infty}\,(|\,xx'\,| + x^2) < \infty.$$

The relation (41) clear holds if $x(t)$ possesses an infinity of zeros (which then necessarily tend to ∞). Thus it can be supposed that $x(t) > 0$ for large t. It can also be supposed that $x(t)$ is monotone for large t, otherwise

the fact that $x(t)$ is of class L^2 implies that $x(t)$ has a sequence of local minima $x(t_n)$ satisfying $x'(t_n) = 0$, $x(t_n) \rightarrow 0$ and $t_n \rightarrow \infty$, as $n \rightarrow \infty$. In which case, (41) holds. But the monotony of $x(t)$ means that

$$\int^t (|xx'| + x^2)ds = |\int^t xx'ds| + \int^t x^2ds = \tfrac{1}{2}x^2 + O(1)$$

has a finite limit inferior, as $t \rightarrow \infty$. Hence (41) holds in any case. This proves (II).

PARIS, FRANCE.

REFERENCES.

[1] G. Borg, "On the point spectrum of $y'' + (\lambda - q(x))y = 0$," *American Journal of Mathemataics*, vol. 73 (1951), pp. 122-126.

[2] P. Hartman and A. Wintner, "The asymptotic arcus variation of solutions of real linear differential equations of second order," *ibid.*, vol. 70 (1948), pp. 1-10.

[3] ——— and A. Wintner, "Criteria of non-degeneracy for the wave equation," *ibid.*, vol. 70 (1948), pp. 295-308.

[4] ——— and A. Wintner, "A criterion for the non-degenracy of the wave equation," *ibid.*, vol. 71 (1949), pp. 206-213.

[5] N. Levison, "Criteria for the limit-point case for second order linear differential operators," *Časopis pro pěstováni matematiky a fysiky*, vol. 74 (1949), pp. 17-20.

[6] D. B. Sears, "Note on the uniqueness of the Green's functions associated with certain differential equations," *Canadian Journal of Mathematics*, vol. 2 (1950), pp. 314-325.

[7] E. C. Titchmarsh, "On the uniqueness of the Green's functions associated with a second order differential equation," *ibid.*, vol. 1 (1949), pp. 191-198.

[8] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[9] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

[10] ———, "On the normalization of the characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

ON BOUNDED GREEN'S KERNELS FOR SECOND ORDER LINEAR ORDINARY DIFFERENTIAL EQUATIONS.*

By Philip Hartman.[1]

1. Let $q = q(t)$, where $0 \leqq t < \infty$, be a real-valued continuous function satisfying

$$q(t) < C^2, \tag{1}$$

where $C > 0$ is a constant. Then the differential equation

$$x'' + (\lambda + q)x = 0 \tag{2}$$

and a homogeneous boundary condition at $t = 0$ determine a self-adjoint boundary value problem in the Hilbert space $L^2 = L^2(0, \infty)$; [10], p. 238, Let the set S of λ-values which are cluster points of the spectrum of such a boundary value problem be called the "essential spectrum" of (2). This terminology is justified by the fact that S is independent of the particular boundary condition which defines it; [10], p. 251.

If λ is a (real) value not in S, then (2) possesses a solution $x = z(t) \not\equiv 0$ of a class L^2; [5]. It has been shown by Wintner [13] that such a solution, $z(t)$, and its derivative, $z'(t)$, are majorized by t^{-n} as $t \to \infty$, for every n, and he has raised the question whether or not the estimate t^{-n} can be replaced by an exponentially small one. This question has been answered in part by Putnam [9], who has shown that such an estimate exists if condition (1) is strengthened to

$$| q(t) | < C^2 \tag{3}$$

and λ is subject to the restriction that $| \lambda + q | \geqq$ const. > 0 for large t. The object of this paper is to answer Wintner's question (without any restrictive hypothesis) in the affirmative.

(*) *If $q = q(t)$ is a continuous function on $0 \leqq t < \infty$ satisfying* (1) *and if λ is a real number not in the essential spectrum of* (2), *then there exists a number $\epsilon = \epsilon(\lambda) > 0$ and a solution $x = z(t)$ of* (2) *satisfying*

$$\limsup_{t \to \infty} (z^2 + z'^2) e^{\epsilon t} < \infty; \tag{4}$$

---

* Received November 27, 1950.

[1] John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

every solution of (2), $x = y(t)$, *linearly independent of* $x = z(t)$ *satisfies*

(5) $$\liminf_{t \to \infty} (y^2 + y'^2) e^{-\epsilon t} > 0.$$

It can be mentioned that (*) is false if the condition (1) is dropped; [7], p. 648. Also, the following converse of (*) is false ([4,], p. 860): If q satisfies (1), or even (3), and if (2) possesses non-trivial solutions $x = z(t)$, $y(t)$ satisfying (4), (5), respectively, for some $\epsilon > 0$, then λ is not in S.

As a corollary of the proof of (*), see (**) in § **10** below, it will be seen that ϵ is arbitrary in (4) and (5) if, for every λ, the non-trivial solutions of (2) have only a finite number of zeros.

2. In the proof of (*), it can obviously be supposed that

(6) $$C > 1.$$

It can also be supposed that $\lambda = 0$ (for the addition of λ to q translates S by $-\lambda$); in this case, (2) reduces to

(7) $$x'' + qx = 0.$$

It is understood that only real-valued solutions of (7) occur below.

Since $\lambda = 0$ is not in S, the differential equation (7) possesses a non-trivial solution $x = z(t)$ of class L^2; see [5]. Let $x = y(t)$ be any solution of (2) which is linearly independent of $x = z(t)$. Then the integral operator with the Green kernel $G(s, t) = G(t, s) = z(s)y(t)$ for $0 \leqq t \leqq s < \infty$ is bounded in L^2; cf. [6]. That is to say, if $g(t)$ is of class L^2 and

(8) $$h = h_g(t) = z(t) \int_0^t y(s) g(s) ds + y(t) \int_t^\infty z(s) g(s) ds,$$

then $h(t)$ is of class L^2 and

(9) $$\int_0^\infty h^2 ds \leqq M \int_0^\infty g^2 ds,$$

where M is a constant independent of g.

If $T > 0$ and $g = g_T(t)$ is defined to be 0 or $z(t)$ according as $0 \leqq t \leqq T$ or $T < t < \infty$, then $h_g(t) = y(t) \int_T^\infty z^2 ds$ for $0 \leqq t \leqq T$. Hence (9) implies

$$(10) \qquad \left(\int_0^T y^2 ds\right)\left(\int_T^\infty z^2 ds\right) \leqq M$$

for $0 < T < \infty$.

3. It is clear from the statement of (*) that, without loss of generality, it can be supposed that

$$(11) \qquad y(0) = 0 \text{ or } y'(0) = 0.$$

If (7) is multiplied by x, and xx'' is written in the form $(xx')' - x'^2$, it is seen ([12], p. 9) that (11) and (1) imply

$$\int_0^T y'^2 ds \leqq y(T)y'(T) + C^2 \int_0^T y^2 ds;$$

while (1) implies that $z'(t)$ is of class L^2, $zz' \to 0$ as $t \to \infty$, and

$$\int_T^\infty z'^2 ds \leqq -z(T)z'(T) + C^2 \int_T^\infty z^2 ds$$

([12], p. 9). The last two inequalities are strengthened if C is replaced by $2C$, hence

$$(12) \qquad 2\int_0^T y^2 ds \geqq Y(T) - y(T)y'(T)/4C^2$$

and

$$(13) \qquad 2\int_T^\infty z^2 ds \geqq Z(T) + z(T)z'(T)/4C^2,$$

where

$$(14) \qquad Y(T) = \int_0^T (y^2 + (y'/2C)^2 ds \text{ and } Z(T) = \int_T^\infty (z^2 + (z'/2C)^2) ds.$$

Since $x = y(t)$, $z(t)$ are linearly independent solutions of (7), their Wronskian is a non-vanishing constant

$$(15) \qquad yz' - y'z = w \neq 0.$$

Hence, by the Schwarz inequality,

$$(16) \qquad -Z'Y' = (z^2 + (z'/2C)^2)(y^2 + (y'/2C)^2) \geqq w^2/4C^2 > 0.$$

Thus, (10), (12), (13) and (16) imply that the product $P_1(T,U)P_2(T,U)$, where

$$(17) \qquad P_1(T,U) = (Y(T) - y(T)y'(T)/4C^2)/Y'(U)$$

and

$$(18) \qquad P_2(T,U) = (Z(T) + z(T)z'(T)/4C^2)/(-Z'(U)),$$

is bounded from above for $0 < T, U < \infty$.

For the purpose of § 9, Appendix, below, it can be mentioned that an upper bound for P_1P_2 is $(M)(4)(4C^2/w^2)$; cf. (10), (12), (13) and (16).

4. The object of the remainder of the proof will be to appraise $P_2(T,U)$ from the above for a suitable choice of $U = U(T)$. To this end, it will be convenient to appraise $P_1(T,U)$ from below. First,

$$y^2(T) - y^2(0) = \int_0^T 2yy'ds \leqq 2C \int_0^T (y^2 + (y'/2C)^2)ds = 2CY(T);$$

so that, for sufficiently large T (since y is not of class L^2),

$$(19) \qquad y^2(T) \leqq 9CY(T)/4 \leqq 3CY(T).$$

Also

$$(20) \qquad y'^2(T) - y'^2(0) = \int_0^T 2y'y''ds = -\int_0^T 2qyy'ds.$$

As the case when (3), rather than (1), is assumed is quite simple from this point onward, the proof of (*) will first be indicated for this case. This will make clear the meaning of the computations to follow. By (3) and (20),

$$(21) \quad (y'(T)/2C)^2 \leqq \text{Const.} + \tfrac{1}{2}\int_0^T |yy'|\,ds \leqq (3C/4)\int_0^T (y^2 + (y'/2C)^2)ds$$

for sufficiently large T (since y is not of class L^2). Thus, (19) and (21) give $Y'(T) \leqq 3CY(T)$ for large T. Also,

$$(22) \qquad |yy'/4C^2| \leqq (y^2 + (y'/2C)^2)/4C.$$

Thus, by (19), (21) and (22), it follows that $P_1(T,T)$ is bounded from below, for large T, by $1/3C - 1/4C > 0$. (This inequality was the reason

for replacing C by $2C$ in the inequality preceding (12).) Consequently, $P_2(T, T)$ is bounded from above. The inequality (22), where y is replaced by z, shows that $Z(T)/(-Z'(T))$ is also bounded from above, say by $1/\epsilon$. The proof can now be completed as in [9], pp. 146-147. The differential inequality for Z implies

$$Z(T) \leqq Ke^{-\epsilon T} \qquad (K = \text{const.}) \tag{23}$$

for $0 \leqq T < \infty$. Since $z(\infty) = 0$ ([12], p. 8),

$$z^2(T) = -\int_T^\infty 2zz'ds \leqq 2CZ(T), \tag{24}$$

(cf. formula proceeding (19)) and since $z'(\infty) = 0$ ([12], p. 13),

$$z'^2(T) \leqq 4C^3Z(T) \tag{25}$$

(cf. (20) and (21)). Hence (4) follows from (23), (24) and (25). Then (5) follows from (15) and (16). This proves (*) for the case that (3) is assumed.

For the purpose of § **9**, below, it can be mentioned that the number $1/\epsilon$ above is $(M)(4)(4C^2/w^2)(1/3C - 1/4C)^{-1} + 1/4C$; cf. the end of § **3** and the three sentences following (22).

5. In the general case, where (1) is assumed, the estimates (21), (25) are not available for the derivatives $y'(T)$, $z'(T)$. A substitute is furnished by a Tauberian type of argument adapted from [1].

LEMMA. *Let $q = q(t)$ satisfy the conditions of* (*), *let $x = x(t) \not\equiv 0$ be a solution of* (7) *and $2/C \leqq T < \infty$. If $x'(T) \neq 0$, let $T^* = T^*(T)$ be defined as the largest (smallest) possible value satisfying both $|T^* - T| \leqq 2/C$ and $xx' \geqq 0$ $(\leqq 0)$ for $T \leqq t \leqq T^*$ $(T^* \leqq t \leqq T)$. Then*

$$x'^2(T) \leqq 4C^2x^2(T^*). \tag{26}$$

Proof of the Lemma. Consider the case $xx' \geqq 0$ for $T \leqq t \leqq T^*$. Since x and x' cannot vanish simultaneously, xx' changes sign at a zero of x. Hence x, and therefore, x' do not change sign on the interval (T, T^*). It can be supposed that $x \geqq 0$ and $x' \geqq 0$ (otherwise x is replaced by $-x$). Thus $0 < x(t) \leqq x(T^*)$ for $T < t \leqq T^*$. By (7) and (1), it is seen that $x''(t) \geqq -C^2x(T^*)$ or $x'(t) \geqq -C^2x(T^*)(t - T) + x'(T)$. If (26) does

not hold, then $x'(t) > -C^2x(T^*)(t-T) + 2Cx(T^*)$. The right-hand side of the last inequality is non-negative for $T \leqq t \leqq T + 2/C$; so that, by definition, $T^* = T + 2/C$. Another integration shows that $x(T^*) > -2x(T^*) + 4x(T^*)$, since $x(T) \geqq 0$. But this leads to the contradiction $x(T^*) > 2x(T^*) > 0$, and proves the Lemma for this case. The proof for the case $xx' \leqq 0$ is similar.

6. It will be convenient to simplify the factor $P_1(T, U)$ of the product P_1P_2. It will be shown that

(27) $\qquad (Y(T)/Y'(U))P_2(T, U)$ is bounded from above

for $0 < T, U < \infty$. To this end, the linear transformation $g \rightarrow h$ in (8) will be reconsidered. For an arbitrary g of class L^2, the function h possesses a derivative

$$h' = z'(t)\int_0^t y(s)g(s)\,ds + y'(t)\int_t^\infty z(s)g(s)\,ds, \tag{28}$$

which is absolutely continuous. The second derivative of h satisfies (almost everywhere) the inhomogeneous differential equation

$$h'' + qh = wg,$$

where w is the Wronskian constant (15). Furthermore, $h(0) = 0$ or $h'(0) = 0$ by (8) or (28) and (11); also, $hh' \rightarrow 0$ as $t \rightarrow \infty$; and so

$$\int_0^\infty h'^2 ds \leqq C^2 \int_0^\infty h^2 ds + |w| \int_0^\infty |gh|\,ds$$

([12], p. 9; cf. the derivation of the inequality preceding (12)). Consequently, (9) implies that the linear transformation $g \rightarrow h'$ is bounded,

$$\int_0^\infty h'^2 ds \leqq (C^2M + |w|M^{\frac{1}{2}}) \int_0^\infty g^2 ds. \tag{29}$$

If the function $g = g_T(t)$ is chosen as in the derivation of (10), it is seen that

$$\left(\int_0^T y'^2 ds\right)\left(\int_T^\infty z^2 ds\right) \leqq C^2M + |w|M^{\frac{1}{2}}. \tag{30}$$

Hence, (27) follows from (10), (30), (13) and (16).

For the purposes of § **10**, Appendix, below, it can be mentioned that an upper bound for the function in (27) is $(2M + |w| M^{3}C^{-2}) \cdot 2 \cdot (4C^2/w^2)$.

7. The Lemma and the inequality (19) show that if $T \geqq 2/C$, then either $y'(T) = 0$ or $(y'(T)/2C)^2 \leqq 3CY(T^*)$. Since $Y(T)$ is increasing and $T^* \leqq T + 2/C$, it follows that $(y'(T)/2C)^2 < 3CY(T + 2/C)$ in either case. Also, by (19) for sufficiently large T, $y^2(T) \leqq 3CY(T) < 3CY(T + 2/C)$. Consequently, $Y(T + 2/C)/Y'(T)$ is bounded below by $1/6C$ for large T-values. Hence, by (27), the function $P_2(T + 2/C, T)$ is bounded from above, say by α.

For the purposes of § **10**, below, it can be mentioned that α can be chosen to be $(2M + |w| M^{3}C^{-2})(2)(4C^2/w^2)(6C)$; cf. the end of § **6**.

The inequality (22), in which y is replaced by z, gives $|zz'/4C^2| \leqq -Z'/4C$. Thus, by the definition of $P_2(T, U)$, it follows that Z satisfies the functional-differential inequality

$$Z(T + 2/C) + Z'(T + 2/C)/4C \leqq -\alpha Z'(T) \tag{31}$$

for $0 < T < \infty$. It will be shown that (31), together with the inequalities $Z \geqq 0$, $Z' \leqq 0$ imply the existence of constants $\epsilon > 0$, K satisfying (23).

Let $0 \leqq U < T$ and let (31) be integrated, with respect to T, over the interval $[U, T]$. One obtains

$$\alpha Z(T) + Z(T + 2/C)/4C \leqq \alpha Z(U) + Z(U + 2/C)/4C - \int_U^T Z(s + 2/C)ds.$$

By the monotony of $Z(T)$,

$$(\alpha + 1/4C)Z(T + 2/C) \leqq (\alpha + 1/4C)Z(U) - \int_U^T Z(s + 2/C)ds.$$

If the abbreviations

$$\alpha + 1/4C = 1/\beta \text{ and } 2/C = \delta \tag{32}$$

are introduced, the last inequality can be written in the form

$$Z(T + \delta) \leqq Z(U) - \beta \int_U^T Z(s + \delta)ds.$$

If $[U, T]$ becomes $[T, T + \delta]$, then $Z(T + 2\delta) \leqq Z(T) - \beta\delta Z(T + 2\delta)$, by virtue of the monotony of Z. The last inequality is equivalent to $Z(T + 2\delta) \leqq Z(T)(1 + \beta\delta)^{-1}$. For a non-negative integer n, let $T = 2n\delta$.

Then $Z(2n\delta + 2\delta) \leqq Z(2n\delta)(1+\beta\delta)^{-1}$; so that, by induction, $Z(2n\delta) \leqq Z(0)(1+\beta\delta)^{-n}$ for $n = 0, 1, \cdots$. Hence, the monotony of Z implies (23) for $0 \leqq T < \infty$, where $K = Z(0)(1+\beta\delta)$ and

$$\epsilon = (1/2\delta)\log(1+\beta\delta) > 0. \tag{33}$$

Consequently, (23) and (24) show that

$$z^2(T) \leqq 2CKe^{-\epsilon T}$$

for $0 \leqq T < \infty$. By the Lemma, if $T = 2/C$, then either $z'(T) = 0$ or (26) holds. But since $T^* \geqq T - 2/C$, one has

$$z'^2(T) \leqq (8C^3Ke^{2\epsilon/C})e^{-\epsilon T}$$

in either case. Hence (4) holds, while (5) is a consequence of (16) and (6). This completes the proof of (*).

8. *Remark.* When (3) is assumed, one can show that, not only does (4) and (5) hold but, also there exists an $\eta > 0$ satisfying

$$\liminf_{t\to\infty} (z^2 + z'^2)e^{\eta t} > 0 \text{ and } \limsup_{t\to\infty} (y^2 + y'^2)e^{-\eta t} < \infty. \tag{34}$$

Thus, for large t,

$$e^{-\eta t} < z^2 + z'^2 < e^{-\epsilon t} \text{ and } e^{\epsilon t} < y^2 + y'^2 < e^{\eta t}$$

for suitably chosen positive ϵ and η.

The relations (34) are known ([11], p. 391) and do not depend on the assumption that λ is not in S. They also are a consequence of the fact that $Y(T)/Y'(T)$ is bounded from below for large T (cf. (19) and (21)).

The relations (34) need not hold if (3) is relaxed to (1); cf. § **10**, below. This is shown by the example in which $q(t) = 2 - 4t^2$. Since $q(t) \to -\infty$, as $t \to \infty$, the set S is empty; [10], p. 238. On the other hand, the case $\lambda = 0$ of (2), that is, (7) has the solution $x = z(t) = \exp(-t^2)$. Hence, the first relation in (34) is violated for every η. In particular, $Y(T)/Y'(T)$ cannot be bounded from below for large T, although if C is chosen to be 2, then $Y(T+1)/Y'(T)$ is (cf. § 7, above).

Appendix.

9. The essential spectrum S is a closed set; so that, its complement is an open set. The intervals on the real λ-axis occurring in the canonical decomposition of this open set will be called the gaps in the essential spectrum of (2). For a given λ not in S, let $L(\lambda)$ denote the least upper bound of the set of ϵ satisfying (4) and (5). (Since $z = z(t)$ is unique up to a constant factor, it is clear that L is determined by λ and not by any choice of solutions of (2).) In a recent paper [4], one of the problems considered was an upper estimate for $L(\lambda)$, as $\lambda \to \infty$ (on the gaps of the essential spectrum, if any) in the case that q satisfies (3). Since the above proof for the existence of an $\epsilon = \epsilon(\lambda)$ is "constructive," it provides a lower estimate for $L(\lambda)$. It can be expected that the problem of obtaining lower estimates for $L(\lambda)$ can be quite delicate for, as suggested by Wintner, if λ is on a gap and tends to an endpoint, then $L(\lambda)$ tends to 0, while $L(\lambda)$ attains its "maximum" on the gap when λ is near the midpoint. Such a behavior is indicated by the lower estimate for $L(\lambda)$ provided by the above proof.

For the case when $\lambda = 0$ and (3) holds, it is seen, by the end of § **4**, that (4) and (5) hold if $\epsilon = (AC^3M + B/C)^{-1}$, where A and B are absolute constants and M is the square of the norm of the Green integral operator (8) (when the Wronskian constant w in (15) is 1).

When λ is arbitrary, C^2 can be replaced by $\lambda + C^2$; so that for large λ, the "C" in the formula for ϵ is asymptotically equal to $\lambda^{\frac{1}{2}}$. Also, $M = M(\lambda)$ is the reciprocal of the square of the distance from λ to the nearest point of the spectrum belonging to the homogeneous boundary condition satisfied by $x = y(t)$ at $t = 0$. If λ is fixed, it is seen from the characterization of the spectrum given in [3] that if the interval $[0, \infty)$ is replaced by $[T, \infty)$ and $T \to \infty$, then $M(\lambda) = M_T(\lambda) \to D^{-2}(\lambda)$, where $D(\lambda)$ is the distance from λ to the nearest point of the essential spectrum. (This limit process clearly does not affect $L(\lambda)$.) Finally, (3) implies that $D(\lambda)$ is bounded, as $\lambda \to \infty$ ([8]; cf. also [4]); so that the term $B/C \sim B/\lambda^{\frac{1}{2}}$ is small when compared to $AC^3M \sim A\lambda^{3/2}D^{-2}(\lambda)$, as $\lambda \to \infty$.

Summarizing, if (3) holds, one obtains the lower estimate Const. $D^2(\lambda)/\lambda^{3/2}$ for $L(\lambda)$ when λ is large (and on a gap of the essential spectrum, if any). This contrasts with the upper estimate Const. $/\lambda^{\frac{1}{2}}$ obtained in [4].

10. In the case that (1) is assumed, it is seen that the ϵ constructed in § **7** is of the form $\epsilon = (C/4) \log \{1 + 8(AC^4M + BC^2M^{\frac{1}{2}} + 1)^{-1}\}$, for the

case when $\lambda = 0$ and C, M have the same significance as in the preceding section, while A, B are absolute constants; cf. (32), (33) and the value for α given in the paragraph preceding (31). It can be remarked that if C^2M is large, as in the preceding section, then ϵ is of the same order of magnitude, $1/C^3M$, as obtained above. But our interest in this section lies in the opposite direction.

As above, if the half-axis $[0,\infty)$ is replaced by $[T,\infty)$, it is seen that, when the essential spectrum S is not empty, then $M = M_T \to D^{-2}(0)$, as $T \to \infty$, where $D(0)$ is the distance from $\lambda = 0$ to the nearest point of S. But if S is empty (and so, $D(0) = \infty$), it is seen that $M_T \to 0$, as $T \to \infty$. Thus, in this case, one can choose, for example, $\epsilon > \frac{1}{2}(C/4)\log(1+8)$. But the constant C can be chosen arbitrarily large (since it is only subject to the restriction (1)). Hence, in (4) and (5), the number ϵ can be chosen arbitrarily large. Since the essential spectrum is empty, whenever (11) is non-oscillatory for every λ ([2], p. 707), the following theorem is proved:

(**) *Let $q = q(t)$, where $0 \leqq t < \infty$, be continuous and satisfy* (1) *for some constant $C > 0$. In addition, let q have the property that for every λ, where $-\infty < \lambda < \infty$, the non-trivial solutions of* (2) *have only a finite number of zeros for $0 \leqq t < \infty$. Then* (2) *has a solution $x = z(t) \not\equiv 0$ satisfying* (4) *for every ϵ; any solution of* (2), *$x = y(t)$, linearly independent of $x = z(t)$, satifies* (5) *for every ϵ.*

This theorem is trivial if (2) is non-oscillatory for every λ by virtue of the condition $q \to -\infty$, as $t \to \infty$; see, for example, the comparison theorem in [7], p. 635. But (2) can be non-oscillatory for every λ when the limit superior of q, as $t \to \infty$, is positive or even $+\infty$; cf. [2], p. 708. On the other hand, if (2) is non-oscillatory for every λ, but q fails to satisfy (1), then (2) need not have a solution which tends to 0 or is even bounded, as $t \to \infty$; [7], p. 648.

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[2] ———, "Differential equations with non-oscillatory eigenfunctions," *ibid.*, vol. 15 (1948), pp. 697-709.

[3] ———, "A characterization of the spectra of one-dimensional wave equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 915-920.

[4] ——— and C. R. Putnam, "The gaps in the essential spectra of wave equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 849-862.

[5] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[6] ——— and A. Wintner, "On the orientation of unilateral spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 309-316.

[7] ——— and A. Wintner, "Oscillatory and non-oscillatory linear differential equations," *ibid.*, vol. 71 (1949), pp. 627-649.

[8] C. R. Putnam, "The cluster spectra of bounded potentials," *ibid.*, vol. 70 (1948), pp. 842-848.

[9] ———, "On isolated eigenfunctions associated with bounded potentials," *ibid.*, vol. 72 (1950), pp. 135-147.

[10] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürliche Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[11] A. Wintner, "The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[12] ———, "(L^2)-connections between the potential and kinetic energies of linear systems," *ibid.*, vol. 69 (1947), pp. 5-13.

[13] ———, "On the smallness of isolated eigenfunctions," *ibid.*, vol. 71 (1949), pp. 603-611.

ON THE EIGENVALUES OF DIFFERENTIAL EQUATIONS.*

By Philip Hartman.[1]

In the differential equation

$$x'' + (\lambda - q)x = 0, \tag{1}$$

let λ be a real parameter and $q = q(t)$, $0 \leqq t < \infty$, a real-valued continuous function satisfying

$$q(t) \to \infty, \text{ as } t \to \infty. \tag{2}$$

Then (1) and a real homogeneous boundary condition

$$x(0) \sin\theta - x'(0) \cos\theta = 0 \tag{3}$$

determine on $L^2(0, \infty)$ a self-adjoint boundary value problem, the spectrum of which is an unbounded sequence of numbers $\lambda_0 < \lambda_1 < \cdots$, and the eigenfunctions belonging to $\lambda = \lambda_n$ have exactly n zeros on $0 < t < \infty$ (Weyl [6]).

It has been proved by Milne [4] that if q possesses a continuous third derivative and if $q' > 0$, $q'' \geqq 0$ and $q'' = o(q'^{4/3})$, as $t \to \infty$, then, as $n \to \infty$,

$$\pi n \sim \int_0^r (\lambda_n - q)^{\frac{1}{2}} dt, \text{ where } r = r_n \tag{4}$$

is the t-value satisfying $q(t) = \lambda_n$. The assumption of the existence of a third derivative, as pointed out by Milne, was unessential to his proof. Titchmarsh [5] gave a simplified proof of (4) retaining all of Milne's conditions, except that concerning the existence of a third derivative. It has been shown by Wintner and the author [3] that (4) holds under the mere assumption that the graph of $q = q(t)$ is convex upwards; for example, if $q(t)$ has a non-negative second derivative. It will be seen below, however, that the assumption of convexity in all of these proof is quite artificial. In fact, convexity implies that the first derivative $q'(t)$ exists, except possibly

* Received October 31, 1950.

[1] John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

for a denumerable set of t-values, and that $q'(t) \geqq \text{const.} > 0$ for large t. This inequality, or even

$$(5) \qquad q'(t)t^3 \to \infty, \text{ as } t \to \infty,$$

will be seen to be sufficient for (4). On the other hand, in the sufficient condition (5), the exponent 3 cannot be replaced by a larger exponent, $3+\epsilon$; in fact, if $q(t)$ satisfies (2) and possesses a continuous positive derivative for which

$$(6) \qquad 0 < \liminf_{t\to\infty} q'(t)t^3 < \infty$$

holds, then (4) need not hold.

The proof of (4) in [3] depends on the method used in [2] to derive an asymptotic formula for the number of zeros on $(0, t)$ of a solution of a differential equation of the type

$$(7) \qquad x'' + Q(t)x = 0,$$

where no parameter λ occurs. It turns out that the asymptotic formula in question is a simple consequence of the following lemma of Wintner and the author, see [1]:

Lemma. *Let $Q(t) > 0$ be a continuous function of bounded variation on $0 \leqq t \leqq T$ and let M be the number of zeros of a solution $x = x(t) \not\equiv 0$ of (7) for $0 < t < T$. Then*

$$(8) \qquad \left| \pi M - \int_0^T Q^{\frac{1}{2}}(s)\,ds \right| \leqq \tfrac{1}{4} \int_0^T Q^{-1}(s) \left| dQ(s) \right| + 2\pi.$$

Correspondingly, with the aid of this lemma, it will be proved that

(i) *the asymptotic formula (4) holds if $q(t)$ is an increasing function satisfying, as $t \to \infty$,*

$$(9) \qquad X(t) = \operatorname*{g.l.b.}_{t \leqq u < v < \infty} \; (q(v) - q(u)) / \int_u^v s^{-3}\,ds \to \infty$$

(for example, if $q(t)$ has a continuous derivative satisfying (5)),

(ii) *but (4) need not hold if $q(t)$ has a continuous derivative satisfying (6).*

Condition (5), or correspondingly (9), is not a condition on the rate of growth of $q(t)$ but a condition on the regularity of growth of the function

$t = r(\lambda)$ which is inverse to $\lambda = q(t)$. Thus, in terms of $r(\lambda)$, condition (5) is

(5 bis) $$r' = o(r^3), \text{ as } \lambda \to \infty \qquad (r' = dr/d\lambda),$$

while (9) is

(9 bis) $$Y(\lambda) = \underset{\lambda \leqq \mu < \nu < \infty}{\text{l. u. b.}} (r(\nu) - r(\mu)) / \int_\mu^\nu r^3(\sigma) d\sigma \to 0,$$

as $\lambda \to \infty$.

Proof of (i). After a suitable translation of the λ-axis, it can be supposed that $q(0) = 0$ (and so $r(0) = 0$). Let $N = N(\lambda)$ be the number of zeros on $0 < t < r(\lambda)$ of a solution $x = x(t) \not\equiv 0$ of (1) and (3). Since the coefficient function, $\lambda - q$, of (6) is negative for $t > r(\lambda)$, it follows that $N(\lambda_n)$ is either n or $n - 1$. Hence, (4) follows if it is proved that, as $\lambda \to \infty$,

(10) $$\pi N(\lambda) \sim \int_0^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt.$$

For a $\mu = \mu(\lambda)$ satisfying $0 < \mu < \lambda$, let $M = M(\mu, \lambda)$ be the number of zeros of $x = x(t)$ on $0 < t < r(\mu)$. If the Lemma is applied to $Q(t) = \lambda - q(t)$ and $T = r(\mu)$, then, since $| dQ | = dq$, it follows that

(11) $$\pi M = \int_0^{r(\mu)} (\lambda - q)^{\frac{1}{2}} dt + O(\log \lambda - \log(\lambda - \mu)) + O(1),$$

where the constant implicit in the O-terms are independent of λ and μ. For later purposes, it should be remarked that the term $O(\log \lambda)$ in (11) is unimportant, since the expression on the right-hand side of (10) obviously exceeds const. $r(\frac{1}{2}\lambda)\lambda^{\frac{1}{2}} >$ const. $\lambda^{\frac{1}{2}}$.

On the interval $r(\mu) \leqq t \leqq r(\lambda)$, the coefficient function $\lambda - q$ of (1) is not greater than $\lambda - \mu$. Hence, by the Sturm comparison theorem, the number, $N - M$, of the zeros of the solution $x = x(t)$ on this interval satisfies

(12) $$\pi(N - M) \leqq (r(\lambda) - r(\mu))(\lambda - \mu)^{\frac{1}{2}} + \pi.$$

It is also clear that

(13) $$\int_{r(\mu)}^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt \leqq (r(\lambda) - r(\mu))(\lambda - \mu)^{\frac{1}{2}}.$$

On the other hand,

$$(14)\qquad \int_0^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt \geqq \int_0^{r(\mu)} (\lambda - q)^{\frac{1}{2}} dt \geqq r(\mu)(\lambda - \mu)^{\frac{1}{2}}.$$

Hence, if $\mu = \mu(\lambda)$ satisfies, as $\lambda \to \infty$,

$$(15)\qquad r(\lambda) \sim r(\mu),$$

the expressions on the left-hand sides of (12) and (13) are equal to an $o(1)$-term times the integral in (10). Thus, by (11) and the remark following it,

$$(16)\qquad N(\lambda) = (1 + o(1)) \int_0^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt + O(|\log(\lambda - \mu)|).$$

Consequently, assertion (i) will be proved if it is shown that $\mu = \mu(\lambda)$ can be chosen so as to satisfy (15) and

$$(17)\qquad \log(\lambda - \mu) = o\Big(\int_0^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt\Big), \text{ as } \lambda \to \infty.$$

To this end, let

$$(18)\qquad \mu = \mu(\lambda) = \lambda - r^{-2}(\lambda);$$

so that, in particular,

$$(19)\qquad \mu \to \infty, \text{ as } \lambda \to \infty.$$

Then, by the definition of Y in (9 bis),

$$r(\lambda) \leqq r(\mu) + Y(\mu) \int_\mu^\lambda r^3(\sigma)\, d\sigma.$$

Hence, by (18) and the monotony of $r(\lambda)$, it follows that $r(\lambda) \leqq r(\mu) + Y(\mu) r(\lambda)$. Thus $1 \leqq r(\lambda)/r(\mu) \leqq (1 - Y(\mu))^{-1}$. Consequently, (15) follows from (9 bis) and (19).

The definition of X in (9) implies that

$$\lambda - q(t) \geqq X(t) \int_t^{r(\lambda)} s^{-3} ds = \tfrac{1}{2} X(t)\,(t^{-2} - r^{-2}(\lambda)).$$

If $0 < t \leqq \frac{1}{2} r(\lambda)$, then $(t^{-2} - r^{-2}(\lambda)) \geqq 3t^{-2}/4$. Consequently, for

$$r^{\frac{1}{2}}(\lambda) \leqq t \leqq \tfrac{1}{2} r(\lambda), \text{ one has } \lambda - q(t) \geqq (3/8) X(r^{\frac{1}{2}}(\lambda))\, t^{-2};$$

so that the integral in (17) exceeds $\frac{1}{2}X^{\frac{1}{2}}(r^{\frac{1}{2}}(\lambda))\log\frac{1}{2}r^{\frac{1}{2}}(\lambda)$. On the other hand, the term on the left-hand side of (17) is $\log r^2(\lambda)$, by (18). Hence, (17) follows from (9). This completes the proof of (i).

Proof of (ii). This proof is similar to the construction of the counter-example in [2]. By an induction (to be indicated below), define an increasing sequence of positive numbers $\alpha_1, \alpha_2, \cdots$ which tend rapidly to ∞. It is then possible to define an increasing, unbounded function $q = q(t)$ which, on the interval $1 + \alpha_k \leqq t \leqq \alpha_{k+1}$, is equal to $k - t^{-2}$ and which has a continuous derivative satisfying $q'(t)t^3 \geqq 2$ for all large t. It will be seen that if α_{k+1} is chosen so that $\alpha_{k+1} - \alpha_k$ is sufficiently large with respect to α_k, then (4) cannot hold.

If a and η are constants, the solutions of $x'' + t^{-2}x = 0$ are of the form $x = x(t) = at^{\frac{1}{2}}\cos((3^{\frac{1}{2}}/2)\log t + \eta)$ and have therefore an infinity of zeros for $1 \leqq t < \infty$. Hence, if $\alpha_{k+1} - \alpha_k$ is sufficiently large, a solution of (1), when $\lambda = k$, has an arbitrarily large number of zeros on the interval $0 < t < \alpha_{k+1}$. Thus, since an eigenfunction belonging to $\lambda = \lambda_n$ has exactly n zeros, it follows that if $\alpha_{k+1} - \alpha_k$ is sufficiently large, the boundary value problem (1), (3) has an eigenvalue $\lambda = \lambda_n$ arbitrarily near, but less than, k. Let $\beta_k = (1 + \alpha_k)\exp(k\alpha_k)$ and let α_{k+1} be so large that there exists an eigenvalue $\lambda = \lambda_n$ satisfying $k - \beta_k^{-2} \leqq \lambda_n < k - \alpha_{k+1}^{-2}$. In particular, $\alpha_{k+1} > r(\lambda_n) \geqq \beta_k$.

In order to appraise the number $N(\lambda_n)$, consider the Sturm majorant of $\lambda_n - q(t)$ which is equal to k for $0 \leqq t \leqq 1 + \alpha_k$ and to t^{-2} for $1 + \alpha_k \leqq t \leqq r(\lambda_n)$. It follows that

$$\pi N(\lambda_n) \leqq k^{\frac{1}{2}}(1 + \alpha_k) + (3^{\frac{1}{2}}/2)\log\{r(\lambda_n)/(1 + \alpha_k)\} + 4\pi.$$

Since $r(\lambda_n) \geqq \beta_n$, it is seen that $\log\{r(\lambda_n)/(1 + \alpha_k)\} \geqq k\alpha_k$. Thus the first term on the right-hand side in the last formula line is negligible, that is,

$$\pi N(\lambda_n) \leqq (1 + o(1))(3^{\frac{1}{2}}/2)\log\{r(\lambda_n)/(1 + \alpha_k)\}.$$

On the other hand,

$$\int_0^{r(\lambda_n)} (\lambda_n - q)^{\frac{1}{2}}dt \geqq \int_{1+\alpha_k}^{r(\lambda_n)} t^{-1}dt = \log\{r(\lambda_n)/(1 + \alpha_k)\}.$$

Consequently,

$$\liminf_{n\to\infty} \pi N(\lambda_n) \Big/ \int_0^{r(\lambda_n)} (\lambda_n - q)^{\frac{1}{2}}dt \leqq 3^{\frac{1}{2}}/2 < 1.$$

This completes the proof of (ii).

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "The number of L^2-solutions of $x'' + q(t)x = 0$," *American Journal of Mathematics*, vol. 73 (1951), pp. 635-645.

[2] ——— and A. Wintner, "The asymptotic arcus variation of solutions of linear differential equations of second order," *ibid.*, vol. 70 (1948), pp. 1-10.

[3] ——— and A. Wintner, "On the asymptotic problems of the zeros in wave mechanics," *ibid.*, vol. 70 (1948), pp. 461-480; see Appendix, pp. 475-480.

[4] W. E. Milne, "On the degree of convergence of expansions in an infinite interval," *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 906-918.

[5] E. C. Titchmarsh, *Eigenfunction Expansions*, Oxford, 1946; see Chap. VII.

[6] H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunctionen," *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 1909, pp. 37-63.

PARTITIONS AND THEIR REPRESENTATIVE GRAPHS.*

By JAMES K. SENIOR.

Part I. Introduction.

The phenomenon of isomerism among organic chemical compounds raises some questions which may be answered in part by the solution of certain problems falling within the scope of the mathematical theory of graphs. In the present paper, however, the chemical aspects of these problems are mentioned only in passing; it is intended to treat them more fully in another publication. Most of the terminology hereafter employed is derived from the book by Dénes König entitled *Theorie der endlichen und unendlichen Graphen* (Leipzig, 1936). But, for the sake of brevity, the unmodified word "graph" is here applied only to those diagrams which are more fully described as *finite undirected* graphs. Such a graph may or may not contain loops, and it may or may not be connected.

Consider a finite set P of finite positive integers $p_1, p_2, \cdots p_n$ which (solely for the sake of convenience) are arranged in non-increasing order so that $p_1 \geqq p_2 \geqq \cdots \geqq p_n$. Every such set of integers may be regarded as a partition of the integer ΣP where $\Sigma P = p_1 + p_2 + \cdots + p_n$. It will therefore hereafter be called a partition, with the understanding that ΣP is the integer of which P is the partition. Note that, although $P_a = P_b$ always implies $\Sigma P_a = \Sigma P_b$, the converse statement is true only in the trivial case where $\Sigma P = 1$.

Consider also a finite set C of points $c_1, c_2, \cdots, c_n$ and a finite set of segments called *edges* drawn between these points. If and only if an edge is terminated at both ends by the same point, that point will be called a *loop-point* and that edge will be called a *loop-edge.* The two together constitute a *loop.* A loop-edge will be said to *depend* from its uniquely-determined loop-point. A loop-edge can be part of only one loop, but a loop-point may have an indefinite number of dependent loop-edges, and may besides be one terminal point of an indefinite number of non-loop edges.

The *valence* (ramification order) of a point which terminates r non-loop edges and s loop-edges is defined as $r + 2s$. For any point, either r or s but not both may be zero. If edges are drawn between the points of C in such a

* Received August 9, 1949; revised September 10, 1950.

way that each point c_i has the valence p_i, then the graph G thus formed is called a *representative of P* or a *graph for P*; it may be indicated by the symbol $G(P)$. The graph G' is considered to be identical with the above graph G if and only if $n = n'$, and the points (vertices) of G' can be so labelled that:

(1) For every pair of points (c_i', c_j') in G', the number of edges joining c_i' directly to c_j' is the same as the number of edges in G joining c_i directly to c_j.

(2) For every point c_i' in G', the number of loop-edges dependent from c_i' is equal to the number of loop-edges dependent from c_i in G.[1]

Every graph uniquely determines the partition which it represents, for that partition is merely an ordered list of the valences of the points (vertices) of the graph; but not every partition is represented by just one graph. That is to say, the graphs G and G' cannot be identical unless they are graphs for the same P; but they may well be graphs for the same P without being identical.

Let $Z(P)$ be the number of distinct graphs for P.
$ZC(P)$ be the number of distinct connected graphs for P.
$ZL(P)$ be the number of distinct loopless graphs for P.
$ZCL(P)$ be the number of distinct connected loopless graphs for P.

Each of these four non-negative integers is uniquely determined by P.

Evidently (a) If $Z(P) = 0$; then $ZC(P) = ZL(P) = ZCL(P) = 0$
(b) If $ZC(P) = 0$ or $ZL(P) = 0$, then $ZCL(P) = 0$.

It It will hereafter be shown that

(c) If and only if $Z(P) > 0$, then $ZC(P) > 0$ or $ZL(P) > 0$
(d) If and only if $ZC(P) > 0$ and $ZL(P) > 0$, then $ZCL(P) > 0$.

But, for the present paper, the question of chief interest is: When is $ZCL(P)$ less than, when it is equal to, and when is it greater than one?

Although every P uniquely determines its $ZCL(P)$, a given non-negative integer taken as $ZCL(P)$ does not usually (in fact, not in any known case)

[1] There must be x ($\geqq 0$) different ways to label the points (vertices) of G' so as to establish the identity of G and G'. The only question here at issue is: Is $x = 0$, in which case G and G' are distinct; or is $x > 0$, in which case G and G' are identical? Otherwise, the value of x is immaterial for the present discussion.

uniquely determine a corresponding P. Hence there arise two converse problems:

(A) For any given P, determine the unique value of $ZCL(P)$.

(B) For any finite non-negative integer taken as $ZCL(P)$, determine the corresponding class of P's.

A complete solution of these problems is probably beyond the reach of present-day mathematics. (At least, the author knows of no publication where even an attempt is made to obtain a solution for either one of the general problems indicated.) But more restricted problems in the same field may, in some instances, be solved.

Problem A. When $ZCL(P) > 0$ and n is small, it is easy to construct one or several distinct connected loopless graphs for P, and to show that no other such graph for P exists. Thus, $ZCL(P)$ is determined for the particular P in question. But when P is expressed in terms of variables each of which is allowed to take the value of any finite positive integer, to determine $ZCL(P)$ as a function of these variables may be a matter of considerable difficulty.[2]

Problem B. It is again easy to find by trial and error, partitions which can be proved to correspond to given (small) values of $ZCL(P)$. But the author is not aware of any previous work dealing with problem B in any more general fashion. In the present paper, the classes of partitions for which $ZCL(P) = 0$ and for which $ZCL(P) = 1$ are determined.[3] These results carry with them the determination of the more general class where $ZCL(P) > 1$. When $ZCL(P) = 0$, P will be called *nullgraphic*; when $ZCL(P) = 1$, P will be called *unigraphic*, and the one representative connected loopless graph will be called *unique*; when $ZCL(P) > 1$, P will be called *multigraphic*.

---

[2] As long ago as 1874, Cayley attempted to obtain the value of $ZCL(P)$ for every partition meeting the following condition: $n = 3x + 2$; $p_1 = p_x = 4$; $p_{x+1} = p_{3x+2} = 1$. He was led to consider this particular class of partitions because, for each member of this class, $ZCL(P)$ is the number of structurally isomeric saturated hydrocarbons of the formula C_xH_{2x+2}. These compounds play an important role in structural organic chemistry. Cayley's papers appear in *Brit. Assoc. Adv. Sci. Reports*, p. 275 (1875); *Phil. Mag.* (4), *47*, 444 (1875); (5), *3*, 37 (1877). They are also reprinted in Volume IX of his collected works. Since the publication of these first papers, there have appeared in the chemical literature numerous articles dealing in part with Cayley's problems and in part with problems closely related to his.

[3] Chemists have strong reasons for paying special attention to these two particular values of $ZCL(P)$; moreover, of all possible values of $ZCL(P)$, these are the two for which it is by far the easiest to determine the corresponding classes of partitions.

THEOREM (1. 1). $\Sigma P = 2x$ *(where x is any positive integer) is a necessary and sufficient condition for* $Z(P) > 0$.

Where $Z(P) > 0$, let $Ft(P) = \Sigma P/2 - (n-1)$, and let $Fr(P) = \Sigma P/2 - p_1$.[4]

THEOREM (1. 2). $Ft(P) \gtreqless 0$ *is a necessary and sufficient condition for* $ZC(P) > 0$.

THEOREM (1. 3). $Fr(P) \gtreqless 0$ *is a necessary and sufficient condition for* $ZL(P) > 0$.

THEOREM (1. 4). $Ft(P) \gtreqless 0 \lesseqgtr Fr(P)$ *(the conjunction of conditions (1. 2) and (1. 3)) is a necessary and sufficient condition for* $ZCL(P) > 0$. *That is to say, if, for any given P, neither the class of connected graphs nor the class of loopless graphs is the null class, then the intersection of these two classes cannot be the null class.*

Hereafter those P's and only those P's which meet condition (1. 4) and hence are represented by at least one connected loopless graph will be called Π's.

THEOREM (2). *Any one of the following conditions (2. 1)-(2. 6″) is sufficient and at least one of them is necessary for* $ZCL(\Pi) = 1$.

UNIQUENESS CONDITIONS.

(2. 1)	$Fr(\Pi) = 0$.	*One half of the sum of all the integers in* Π *is equal to the largest integer in* Π.
(2. 2)	$n = 3$.	*The number of integers in* Π *is 3.*
(2. 3)	$p_3 = 1$.	Π *contains at least three integers but not more than two of them are* > 1.
(2. 4)	$p_1 = 2$.	*The largest integer in* Π *is 2.*
(2. 5)	$Fr(\Pi) = 1;\ p_2 = p_n$.	*The largest integer in* Π *is one less than half the sum of all the integers; excepting this largest integer, all the integers in* Π *are equal.*
(2. 6)	$p_1 = p_3;\ p_4 = 1$.	*The three largest integers in* Π *are all equal to one another; the fourth integer is 1.*
(2. 6′)	$n = 4$.	*The number of integers in* Π *is four.*
(2. 6″)	$Ft(\Pi) = 0$.	*The number of integers in* Π *is* $\Sigma\Pi/2 + 1$.

[4] As is more fully explained in Part II of this paper, the designations $Ft(P)$ and $Fr(P)$ arise as follows. If $Ft(P) = 0$, then any connected graph for P is a *tree*; if $Fr(P) = 0$, then any loopless graph for P is a *rosette*.

That is to say, a Π is unigraphic if it meets any one of the above conditions; it is multigraphic if it meets no one of them.[5]

So far, it has been assumed that any two points in C, at least until they have been incorporated in a graph and sometimes even then, are indistinguishable from one another. Consider now in C a set I consisting of $s_i > 1$ points to all of which the same valence p_i is to be assigned. Break I up into disjoint subsets which together exhaust I. These subsets may be made distinguishable from one another by endowing all the points of each particular subset with an intrinsic property such as one particular color, and assigning a different color to each subset. A graph constructed with a set C of points which contains at least one parti-colored set I will be called a *polychromatic* graph to distinguish it from the *monochromatic* graphs hitherto considered.[6]

The identity of two polychromatic graphs is fixed as follows: If G and G' are identical, the corresponding points c_i in G and c_i' in G' must not only be of like valence and similarly attached by edges to themselves and to the other points of G and G' respectively (the purely topological condition which holds for the monochromatic graphs hitherto discussed), but, in addition, the like-valent points c_i' and c_j' in G' must be alike in color if and only if the corresponding like-valent points c_i and c_j in G are alike in color.

THEOREM (3). *The uniqueness conditions for connected loopless graphs must be successively strengthened as the number of colors in these graphs is*

[5] The classes of partitions for which $ZCL(P) = 0$ and for which $ZCL(\Pi) = 1$ are both infinite. Moreover the conditions for $ZCL(\Pi) = 1$ are much more complex than those for $ZCL(P) = 0$. It appears probable (though it is not here proved) that the conditions for $ZCL(\Pi) = x$ become increasingly intricate as x increases. Probably, for every positive value of x, the class of corresponding Π's is infinite. But the author knows of no proof that such is the fact; indeed, there is no proof that, for at least one value of x, the class of corresponding partitions may not be the null class.

For structural organic chemistry, no partition which is not a Π has any significance, and all Π's are subject to the mathematically arbitrary but chemically important limitation $p_1 < 5$. When this limitation is imposed, the class of unigraphic partitions shrinks to two infinite sequences (where $p_1 = 2$) plus about thirty individually defined members. Since no member of either of the two infinite sequences has any chemical importance when $n > 4$, the entire number of unigraphic partitions which might possibly be of chemical significance shrinks to less than 35. To examine the chemical representatives (if any exist) of these partitions in order to see whether they possess the proper chemical requirements, is a brief task.

[6] So far as the number of distinct graphs for a given P is concerned, polychromism among points of unlike valence is a vacuous phenemenon. This assertion follows from the fact that, in matching the points of two graphs G and G' to see whether these graphs are or are not distinct, no two points of unlike valence can ever correspond to one another, whether they are of the same color or not.

increased from one to four; a unique connected loopless graph with more than three colors is a rosette.

Of the theorems so far given, (1. 1)-(1. 4) are proved in part II, (2) is proved in part III, and (3) is proved in part IV of the present paper. In each instance, the proofs cannot well be given until certain methods of constructing and classifying graphs have been described. The necessary definitions and descriptions therefore precede the actual proofs in each of the three following parts.

Part II. Proof of theorems (1. 1)-(1. 4).

Symbolism and terminology. If, in the graph G, any part H is a connected graph, and if (excepting H itself) no part of G which includes H is a connected graph, then H is a *component* of G. The number of distinct components which make up the whole of G is $\nu(G)$, and where G is a finite graph, $\nu(G)$ is a positive finite integer. The expression $\nu(G)=1$ means that G itself is connected. If and only if, in G, at least one edge is terminated by both points c_i and c_j, then these two points form in G an *attached* pair. This term 'attached' is far stronger than the term 'connected' which, when applied to a pair of points, refers to any two points in one and the same component of G.

The set of all those edges which attach c_i to c_j is called the ***bond*** (i, j). The number of edges in the bond (i, j) is called the ***breadth*** of that bond and is indicated by the symbol $b(i, j)$. When $b(i, j)=1$, the bond is *single*; when $b(i, j)>1$, the bond is *multiple.* Two points c_i and c_j together with the q edges which form the bond between them may be indicated in a diagram by $c_i \overset{q}{\text{——}} c_j$ and symbolically by (c_i, q, c_j). This symbol implies and is implied by the equation $b(i, j)=q$. The scheme just described may be extended to cover the case (c_i, q, c_i) or $b(i, i)=q$ where q is the number of loop-edges dependent from c_i. Any graph for P is uniquely determined if the value is given for every $b(i, j)$ where $i=1, \cdots, n$ and $j=1, \cdots, n$. In all cases, $b(i, j)=b(j, i)$, and $p_a=\sum_{i=n}^{i=1} b(a, i)+b(a, a)$. Where, for a particular pair of points (c_i, c_j), $b(i, j)=0$, that fact will usually be indicated by omitting the equation and the corresponding symbol $(c_i, 0, c_j)$; but occasionally it is convenient to use such an equation or such a symbol explicitly. Particularly, when one equation or one symbol is used to describe the bond (i, j) in various different graphs, then $b(i, j)$ may be expressed in terms of variables. For example, in the equation $b(i, j)=q-x$ and in the symbol

$(c_i, q-x, c_j)$, let q be a constant and x a variable. Both are positive integers. The scheme just outlined permits $x = q$ as a special case, but under no circumstances may $b(i,j)$ be negative. This same symbolic scheme may be still further extended to apply to subgraphs containing more than two points. The meanings of the symbols $(c_i, b(i,j), c_j, b(j,k), c_k)$ and $(c_i, b(i,j), c_j, \cdots, c_u, b(u,v), c_v)$ should be self-evident. Such a symbol corresponds to a set of equations, not to any single equation.

If $b(i,j) = p_i$, then c_i is a *blind point*. If, in addition, $b(i,j) = 1$, then c_i is an *end point*, and the single edge attaching c_i to c_j is an *end edge*. If, in any connected loopless graph, the number of blind points exceeds $n-2$, the graph is a *rosette*. If all the blind points in a rosette are like-valent, the rosette is *regular*; otherwise it is *irregular*. A regular rosette in which all the blind points are end points is a *star*.

Rosettes are an important and striking class of connected loopless graphs. They are characterized by the property $Fr(\Pi) = 0$, and have the following other properties:

(1) According to condition (1.4), rosettes border the field of connected loopless graphs on one side.

(2) According to condition (2.1), every rosette is unique.

(3) If any Π is represented by a rosette, it cannot be represented by any other loopless graph, connected or disconnected.

If c_1 terminates every non-loop edge in a graph and is moreover the only loop-point in that graph, then the graph will be called a *loop-rosette*. This definition includes as special cases those connected graphs in which the number of non-loop edges is 0.

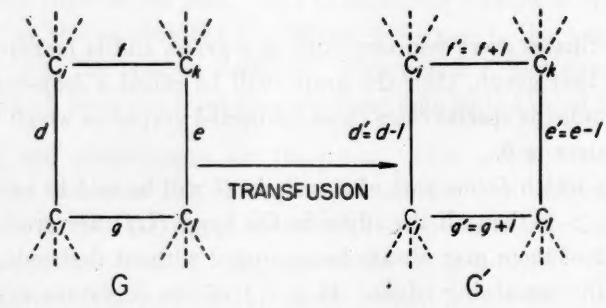
Any edge which forms part of a cycle in G will be said to be *cyclic* in G. If $b(i,j) = q > 1$, then all the edges in the bond (i,j) are cyclic in G, and at most $q-2$ of them may always be expunged without destroying the cyclic character of the remaining edges. If $q-1$ of the edges are expunged, the one remaining edge is non-cyclic *unless*, in G, c_i and c_j are both members of some one and the same cycle of order > 2. A connected graph which has no cyclic edge is a *tree*.

Since every graph uniquely defines its partition, the expressions $Ft(G)$ and $Fr(G)$ may legitimately be used. Furthermore, if G and G' both represent the same P, then $Ft(G) = Ft(G')$ and $Fr(G) = Fr(G')$. For a tree, $Ft(G) = Ft(P) = \Sigma P/2 - (n-1) = 0$; for a rosette, $Fr(G) = Fr(P) = \Sigma P/2 - p_1 = 0$.

Transfusion. Consider the graph G containing the points c_i, c_j, c_k and c_l. Let $b(i,j) = d$, $b(k,l) = e$, $b(i,k) = f$, $b(j,l) = g$. Decrease by one the breadths of the bonds (i,j) and (k,l) by expunging one edge from each of these bonds; increase by one the breadths of the bonds (i,k) and (j,l) by adding one edge to each of these bonds. By this process there is formed the new graph G' where $d' = d - 1$, $e' = e - 1$, $f' = f + 1$, $g' = g + 1$. It is *essential* for the carrying out of this process that, in G, $d > 0 < e$—i. e., that in G the pairs of points (c_i, c_j) and (c_k, c_l) be attached pairs. It is *inevitable* that, after the carrying out of the process, $f' > 0 < g'$—i. e., that in G' the pairs of points (c_i, c_k) and (c_j, c_l) be attached pairs. Hence the process just described is called the *transfusion* of the bond (i,j) with the bond (k,l).[7] As for the bonds (i,k) and (j,l) in G as well as the bonds (i,j) and (k,l) in G', their breadths may be equal to or > 0, depending on the particular G on which the transfusion is performed. (Any bonds in G and G' other than the four here mentioned are common to both of these graphs; they are not involved in the transfusion in question and are irrelevant to that process.) Symbolically the transfusion just described may be expressed by the following equations:

$$b'(i,j) = b(i,j) - 1, \qquad b'(i,k) = b(i,k) + 1,$$
$$b'(k,l) = b(k,l) - 1, \qquad b'(j,l) = b(j,l) + 1.$$

Diagrammatically expressed, it is as follows:



In a transfusion, no one of the points involved undergoes any change in valence, and hence, *P is invariant under transfusion.* In the three possible cases where a transfusion decreases, leaves invariant or increases the number

[7] Note that when a transfusion is employed in a construction, the pairs of points (i,j) and (k,l) are to be regarded as ordered pairs. The instruction "Transfuse (i,j) with (k,l)" means that the new edges to be introduced are (i,k) and (j,l). On the other hand, the instruction "Transfuse (i,j) with (l,k)" means that the new edges to be introduced are (i,l) and (j,k). The edges to be expunged are (i,j) and (k,l) in both instances.

of components, the transfusion will be called respectively *connective, neutral* or *disconnective.* Various special cases demand attention.

I. The two bonds (i, j) and (k, l) belong to different components of G. Here a transfusion may be connective or neutral but cannot be disconnective. *The transfusion is connective if and only if at least one of the two expunged edges is cyclic in G.* Where this condition is met $\nu(G') = \nu(G) - 1$.

II. The two bonds (i, j) and (k, l) belong to one and the same component H of G. Here a transfusion may be neutral or disconnective but cannot be connective. The transfusion is disconnective if and only if the two expunged edges form in H two single bonds which constitute the sole connection between that part of H which contains c_i and c_k and that part of H which contains c_j and c_l. Where this condition is met, $\nu(G') = \nu(G) + 1$. For the present discussion, the most interesting transfusions within one component of G are those where at least one of the expunged edges is a loop-edge. Any such process will hereafter be called a *loop-transfusion.*

III. For purposes of transfusion, a loop-point may be regarded as self-attached. Two varieties of loop-transfusion may be recognized:

(a) The two expunged edges are loop-edges dependent respectively from the two distinct loop-points c_k and c_m.

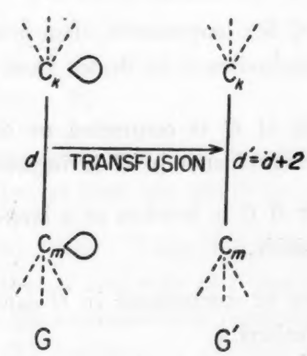
$$b'(m, m) = b(m, m) - 1, \; b'(k, k) = b(k, k) - 1, \; b'(k, m) = b(k, m) + 2.$$

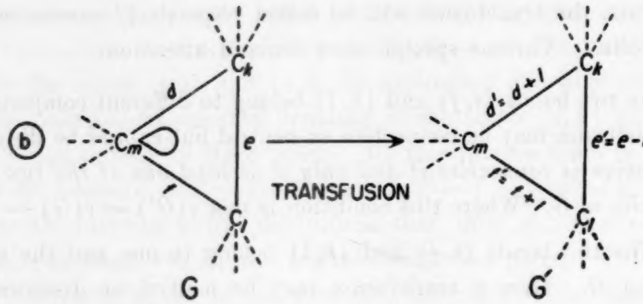
(b) One of the two expunged edges is a loop-edge dependent from the loop-point c_m; the other is an edge in the bond (k, l).

$$b'(m, m) = b(m, m) - 1, \qquad b'(k, m) = b(k, m) + 1,$$
$$b'(k, l) = b(k, l) - 1, \qquad b'(l, m) = b(l, m) + 1.$$

When diagrammatically expressed, these two processes are as follows:

(a)





In diagram (a), d may be equal to or > 0, but d' is necessarily > 0; in diagram (b), d and/or f and/or e' may be equal to or > 0, but e, d', and f' are necessarily > 0.

Since a loop-edge is always cyclic, and since it can belong to no cycle of order > 1, a loop-transfusion is connective if the two expunged edges belong to different components of G; it is neutral if they belong to the same component of G. *Under no circumstances can a loop-transfusion be disconnective.* Process (a) can be carried out if and only if G contains at least two distinct loop-points; it always decreases the number of loops by two. Process (b) always decreases the number of loops by one; but this process is subject to an important limitation. It can be carried out if and only if c_k and c_l constitute an attached pair each member of which is distinct from loop-point c_m. This last condition can always be met if G has at least one component H_i which includes more than one point, and which is distinct from the component H_j containing c_m. But if every non-loop edge in G terminates in c_m, process (b) fails. Hence, if and only if,

(1) c_m is the only loop-point in G, and

(2) every non-loop edge in G terminates in c_m,

that is, if and only if G is a loop-rosette, then processes (a) and (b) both fail.

The important conclusions to be drawn from the above considerations are:

(1) If and only if G is connected or G contains no cyclic edge, a connective transfusion in G is impossible.

(2) If and only if G is loopless or a loop-rosette, a loop-transfusion in G is impossible.

(3) The number of components in G cannot be increased by sucessive loop-transfusions.

(4) The number of loops in G cannot be increased by successive connective transfusions.

Although the object of the procedures hereafter described is to construct loopless connected graphs, it is sometimes convenient, as an intermediate step in such a construction, to reverse a loop-transfusion of type (b). Such a procedure will be called a *reversion,* and when ν is invariant under a particular reversion, that reversion will be called a *neutral reversion.* A neutral reversion can always be performed on a graph G provided that G contains a subgraph defined by $b(m,k) > 1$, $b(m,l) > 0$. Where this condition is met, it is possible, by a neutral reversion, to form a graph G' which contains the subgraph defined by $b(m,m) = 1$, $b(m,k) > 0$, $b(k,l) > 0$. Here $\nu(G) = \nu(G')$.

Proof of Theorem (1.1). Theorem (1.1) states that $\Sigma P = 2x$ (where x is any positive integer) is a necessary and sufficient condition for $Z(P) > 0$. The necessity of condition (1.1) follows directly from the fact that every edge has just two ends. The sufficiency of this same condition is shown by the following construction.

Divide P into two partitions Pe and Po, the first consisting of all the even integers $e_1, e_2, \cdots, e_k$ in P, the second consisting of all the odd integers $o_1, o_2, \cdots, o_{n-k}$ in P. According to condition (1.1), $(n-k)$ is even. Make each point ce_i the loop-point of $e_i/2$ loops; make each point co_i the loop-point of $(o_i - 1)/2$ loops; then attach each point co_{2i} to the point co_{2i-1} by a single edge. By this procedure, a graph for P is formed and the sufficiency of condition (1.1) for $Z(P) > 0$ is proved.

A consequence of Theorem (1.1) is the following

LEMMA. *If all the integers of P_1 are divided into $y-1$ disjoint sets which constitute respectively the partitions $P_2, \cdots, P_y$, and if all but one of these partitions $P_1, \cdots, P_y$ are represented by graphs, then there must always be at least one graph for the remaining y-th partition. This rule is particularly useful where each of the disjoint sets of integers is composed of integers which are consecutive in P_1.*

In all the following discussion, it will be assumed that ΣP is an even positive integer, and hence that at least one graph for P exists.

Proof of Theorems (1.2)-(1.4). Theorems (1.2)-(1.4) are existence theorems. Each one states that if and only if P meets a certain condition, at least one graph for P has the corresponding property. They do *not* imply that if and only if P meets the condition in question every graph for P has

that property.[8] Consequently, the three following proofs are similar in outline. A general method is given for constructing for P a graph with the desired property. It is then proved that the given method fails if and only if P does not meet the corresponding condition.

(A) Theorem (1.2) states that $\Sigma P/2 - (n-1) = Ft(P) \gtreqless 0$ is a necessary and stufficient condition for $ZC(P) > 0$. For the proof of this theorem, further consideration of the function $Ft(P)$ is necessary.

$Ft(P)$ is closely related to but not identical with the well-known graph function μ, called the cyclomatic number of a graph. The function μ is usually defined by $\mu = \alpha_1 - \alpha_0 + \nu$ where α_1 is the number of edges, α_0 is the number of points (vertices), and ν is the number of components of the graph. In the notation here used, $\alpha_1 = \Sigma P/2$ and $\alpha_0 = n$; hence $\mu = \Sigma P/2 - n + \nu$. But ν (although it is a graph function) is (usually) not uniquely determined by P. Moreover, $\mu(G)$ is specially designed so that $\mu(G) < 0$ is an impossibility, whereas $Ft(P)$ is useful chiefly because it sometimes has a negative value. The exact relation between $Ft(G)$ and $\mu(G)$ is as follows:

$$Ft(G) = \Sigma P/2 - n + 1, \qquad \mu(G) = \Sigma P/2 - n + \nu.$$

$$\text{Therefore } Ft(G) = \mu(G) - (\nu - 1).$$

That is to say: $Ft(G) = \mu(G)$ when G is connected, $(\nu = 1)$.

$Ft(G) < \mu(G)$ when G is not connected, $(\nu > 1)$.

It is well known that, where G consists of ν components $H_1, H_2, \cdots, H_\nu$,

(a) if $\nu = 1$ and $\mu(G) = 0$, then G is a tree;

(b) $\mu(G) = \mu(H_1) + \mu(H_2) + \cdots + \mu(H_\nu)$.

Construct for P a graph G_0 which may or may not contains loops and may or may not be connected. Next perform on G_0 a connective transfusion to form G_1. This latter step must always be possible except in the two following instances:

(1) $\nu_0 = 1$. Here $Ft(P) = Ft(G_0) = \mu(G_0) - (\nu_0 - 1)$ reduces to $Ft(P) = \mu(G_0)$; and therefore $Ft(P) \gtreqless 0$.

[8] Evidently, in the highly special cases where $Z(P) = 1$, an existence proof for a particular kind of graph for P becomes a universal proof. Hence the following table.

p_1	n	$\Sigma P/2$	$Ft(P)$	$Fr(P)$	The only possible graph is:
4	1	2	2	-2	connected but not loopless.
1	4	2	-1	1	loopless but not connected.
1	2	1	0	0	loopless and connected.

(2) $\nu_0 > 1$, but G_0 contains no edge cyclic in G_0. Here $H_1, H_2, \cdots, H_\nu$ (the components of G_0) must all be trees. Each $\mu(H_i) = 0$ and hence $\mu(G_0) = 0$. Here $Ft(P) = Ft(G_0) = \mu G_0 - (\nu_0 - 1)$ reduces to $Ft(P) = 1 - \nu_0$; and, since $\nu_0 > 1$, $Ft(P) < 0$. Here $1 - Ft(P) = \nu_0$, the number of trees in G_0.

If $\nu_0 > 1$, and G_0 contains at least one edge cyclic in G_0, then the graph G_1 formed by a connective transfusion performed on G_0 contains $\nu_0 - 1$ components. Exactly the same argument used for G_0 may now be used for G_1. This argument shows that G_1 is connected $(Ft(P) \geqq 0)$, or that G_1 consists of $\nu_1 = (\nu_0 - 1) > 1$ trees $(Ft(P) < 0)$, or that a connective transfusion may be performed on G_1, by which transfusion G_1 is converted into G_2 with $\nu_0 - 2$ components. Successive connective transfusions performed on G_0, G_1, etc., must therefore lead to one of two results:

(a) If $\mu(G_0) \geqq (\nu_0 - 1)$, then $Ft(P) \geqq 0$ and $G_{\nu_0 - 1}$ is a connected graph.

(b) If $\mu(G_0) < (\nu_0 - 1)$, then $Ft(P) < 0$ and $G_{\mu(G_0)}$ is a set of $1 - Ft(P)$ trees.

It is thus proved that $Ft(P) = 0$ is a necessary and sufficient condition for $ZC(P) > 0$.

(B) Theorem (1.3) states that $\Sigma P/2 - p_1 = Fr(P) \geqq 0$ is a necessary and sufficient condition for $ZL(P) > 0$.

Construct for P a graph G_0 which may or may not contain loops and may or may not be connected. Next perform on G_0 a loop-transfusion to form G_1. This latter step must always be possible except in the two following instances:

(1) G_0 is loopless. Here $Fr(G_0) = Fr(P)$ is the number of the edges in G_0 which are *not* terminated by c_1. Evidently, this number cannot be negative, and hence $Fr(P) \geqq 0$.

(2) G_0 is a loop-rosette. Here $p_1 = p_2 + p_3 + \cdots + p_n + 2s$ where $s\ (> 0)$ is the number of loop-edges dependent from c_1. Hence

$$2p_1 > (p_1 + p_2 + \cdots + p_n) = \Sigma P; \quad p_1 > \Sigma P/2;$$
$$\text{and } Fr(P) = (\Sigma P/2 - p_1) < 0.$$

If G_0 is neither loopless nor a loop-rosette, then the number of loops in G_1 must be less than the number of loops in G_0. Exactly the same argu-

ment used for G_0 may now be used for G_1. This argument shows that G_1 is loopless ($Fr(P) \geqq 0$) or that G_1 is a loop-rosette ($Fr(P) < 0$) or that a loop-transfusion may be performed on G_1, by which transfusion G_1 is converted into G_2 which contains less loops than are contained in G_1. Successive loop-transfusions performed in G_0, G_1, etc., must therefore lead to one of two results:

(a) G_i is loopless, and here $Fr(P) \geqq 0$.

(b) G_i is a loop-rosette, and here $Fr(P) < 0$.

That is to say, the only condition which forbids the construction of a loopless graph for P is the only condition which permits $Fr(P)$ to have a negative value. Thus it is proved that $Fr(P) \geqq 0$ is a necessary and sufficient condition for $ZL(P) > 0$.

(C) Theorem (1.4) states that $Ft(P) \geqq 0 \leqq Fr(P)$ (the conjunction of conditions (1.2) and (1.3)) is a necessary and sufficient condition for $ZCL(P) > 0$. Assume that $Ft(P) \geqq 0 \leqq Fr(P)$. Construct a graph G for P. If G is not connected, convert it by successive connective transfusions into a connected graph G' for P. This procedure must be possible since it has been proved to fail if and only if $Ft(P) < 0$. Next, if G' is not loopless, convert it by successive loop-transfusions into a loopless graph G'' for P. This procedure must be possible since it has been proved to fail if and only if $Fr(P) < 0$. The graph G'' thus finally obtained must be connected and loopless because G' is connected and (as has already been shown) no loop-transfusion can be disconnective. It would evidently be impossible to carry out the full procedure just described if either $Ft(P)$ or $Fr(P)$ were negative. Hence $Ft(P) \geqq 0 \leqq Fr(P)$ is a necessary and sufficient condition for $ZCL(P) > 0$.[9]

So far it has been shown that any graph for P may be converted by transfusions into

(1) a connected loopless graph: $Ft(P) \geqq 0$; $Fr(P) \geqq 0$; or

(2) a set of ν (> 1) trees—loopless but not connected: $Ft(P) < 0$; $Fr(P) \geqq 0$; or

(3) a loop-rosette—connected but not loopless: $Ft(P) \geqq 0$; $Fr(P) < 0$.

---

[9] It is of course possible to reverse the order of the transfusions used in the above proof. That is, G may first be converted by loop-transfusions into a loopless graph, and then (since no connective transfusion can increase the number of loops in a qraph) by connective transfusions into a connected loopless graph. The result is the same, whichever method is adopted.

The only other logical possibility is $Ft(P) < 0$; $Fr(P) < 0$. But here $Ft(P) + Fr(P) < 0$; that is,

$$\Sigma(P)/2 - (n-1) + \Sigma(P)/2 - p_1 = \Sigma(P) - (n-1) - p_1 < 0;$$
$$p_2 + p_3 + \cdots + p_n < n-1;$$

and this inequality is impossible if $p_2, p_3, \cdots, p_n$ are all positive integers. That is to say, there is no P which is represented only by disconnected graphs with loops.[10]

Part III. Proof of theorem (2).

(A) The sufficiency of each one of the conditions (2.1)-(2.6″) for $ZCL(\Pi) = 1$.

Each of the uniqueness conditions (2.1)-(2.6″) (p. 666) is independent of all the others. This assertion is proved by the following Π's each one of which fulfills that condition and only that condition to which it corresponds in the following list.

Condition.	Π.	Condition.	Π.
(2.1)	3, 3	(2.5)	4, 2, 2, 2
(2.2)	4, 4, 4	(2.6′)	3, 3, 3, 1
(2.3)	3, 3, 1, 1	(2.6″)	3, 3, 3, 1, 1, 1, 1, 1
(2.4)	2, 2, 2, 2		

But, as thus stated, the uniqueness conditions partially overlap (e.g. the partition 2, 1, 1 fulfills each of the conditions (2.1)-(2.4)), and hence they cannot without some redundancy be proved to be individually sufficient. In such a demonstration it is better to be guided by the following rubric in which the same uniqueness conditions are restated in a form which prevents overlapping.

RUBRIC.

$Fr(\Pi) = 0$. Class I.

$Ft(\Pi) > 0$.

$n = 3$. Class II.

$n > 3$.

[10] One inference to be drawn from the fact that $Ft(P) + Fr(P) \geqq 0$ is that the value of $Ft(P)$ sets a lower limit to the value of $Fr(P)$ and vice versa. There is, in fact, a quasi-dualism between $Ft(P)$ and $Fr(P)$ which is associated with the well-known dualism among partitions. If P_d is the partition dual to P, then $\Sigma P = \Sigma P_d$; $n = p_{1_d}$; $p_1 = n_d$; and hence $Ft(P) + Fr(P) = Ft(P_d) + Fr(P_d)$; $Ft(P) = Fr(P_d) + 1$; $Fr(P) = Ft(P_d) - 1$. This quasi-dualism does not, however, lead to any precise dualism between rosettes and trees.

$p_3 = 1.$ Class III.

$p_3 > 1.$

$p_1 = 2.$ Class IV.

$p_1 > 2.$

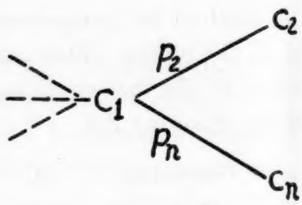
$Fr(\Pi) = 1; p_2 = p_n.$ Class V.

$Fr(\Pi) > 1; p_1 = p_3; p_4 = 1.$

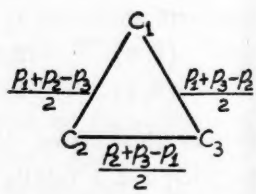
$n = 4.$ Class VI.

$Ft(\Pi) = 0.$ Class VII.

The general graphs for the Π's of classes I-VII are given below. All the Π's of class IV fall into two disjoint subclasses: IVa, where $p_n = 2$, and IVb, where $p_n = 1$.

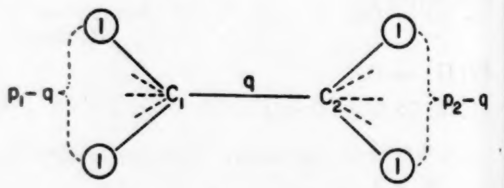


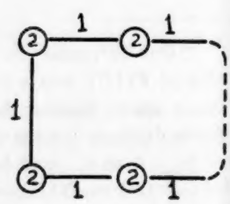
Class I. $b(1, i) = p_i$;
G is a rosette. $i = 2, \cdots, n.$



Class II. $b(1, 2) = (p_1 + p_2 - p_3)/2$;
$b(2, 3) = (p_2 + p_3 - p_1)/2$;
$b(3, 1) = (p_1 + p_3 - p_2)/2.$

Class III. $b(1, 2) = q; 0 < q < p_2$;
$b(1, i) = 1; i = 3, \cdots, (p_1 - q + 2)$;
$b(2, j) = 1; j = (p_1 - q + 3), \cdots, n.$

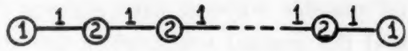




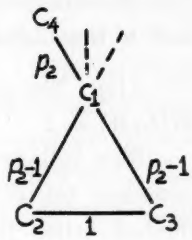
Class IVa. $b(i, i+1) = 1; i = 1, \cdots, (n-1)$;
$b(n, 1) = 1.$

Class IVb. $b(1, n) = 1;$

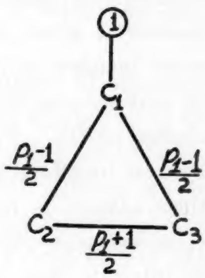
G is a tree. $b(i, i+1) = 1; i = 1, \cdots, (n-2).$



Class V. $b(1,2) = b(1,3) = (p_2 - 1);$
$b(2,3) = 1;$
$b(1,i) = p_2; i = 4, \cdots, n.$

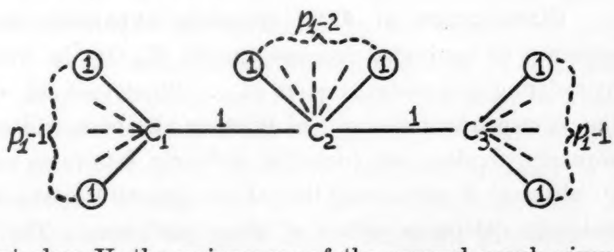


Class VI. $b(1,2) = b(1,3) = (p_1 - 1)/2;$
$b(2,3) = (p_1 + 1)/2;$
$b(1,4) = 1.$



Class VII. $b(1,2) = b(2,3) = 1.$

G is a tree. $b(1,i) = 1; i = 4, \cdots, (p_1 + 2);$
$b(2,j) = 1; j = (p_1 + 3), \cdots, 2p_1;$
$b(3,k) = 1; k = (2p_1 + 1), \cdots, n.$



In every case except class V, the uniqueness of the general graph given for the Π's in question is obvious. The uniqueness of the general graph given for any Π of class V is made evident by the following considerations. Every graph for a Π of class V may be constructed as follows. Where $Fr(\Pi) = 0$, $n > 3$ and $p_2 = p_n > 1$, construct a graph for Π. This graph is a rosette and therefore unique (Class I); it is moreover a regular rosette. In this rosette, expunge two edges so as to decrease $b(1,i)$ and $b(1,j)$ each from p_2

to $(p_2 - 1)$; insert a new edge so as to increase $b(i, j)$ from 0 to 1. A graph for a Π off class V is thus formed. Since all the pairs of points (c_i, c_j) which can be chosen from $c_2, c_3, \cdots, c_n$ are equivalent in the regular rosette, all possible ways of carrying out the indicated construction are equivalent, and the general graph given for any Π of class V is unique.

The sufficiency of each one of the conditions (2.1)-(2.6″) for $ZCL(\Pi) = 1$ is thus demonstrated.

(B) The necessity of at least one of the conditions (2.1)-(2.6″) for $ZCL(\Pi) = 1$.

The full proof that at least one of the conditions (2.1)-(2.6″) is necessary for $ZCL(\Pi) = 1$ is extremely laborious. For this purpose, no method better than the following has so far been found. The Π's are exhaustively classified into disjoint classes. For each of these classes, a method is given for constructing two distinct connected loopless graphs for every member of the class which is in no one of Classes I-VII. The success of this scheme depends on finding a suitable way of defining the disjoint classes of Π's. For, if the number of these classes is too large, the proofs become insufferably repetitious; and if, on the other hand, the number of these classes is too small, then, in some instances, the general method of construction applicable to all the Π's in the class under consideration becomes so intricate that it cannot readily be grasped.

A general principle for classifying Π's is therefore next given. Then follow the definitions of certain concepts which serve to abbreviate the arguments in the ensuing proof.

Classification of Π's; complete expansion sequences. Consider a sequence of connected loopless graphs G_0, G_1—in which each G_i is formed by deleting one cyclic edge in G_{i-1}. Since each G_i uniquely defines its Π_i, the corresponding sequence of Π's may also be considered. In such a *deletion sequence*, G_i does not (usually) uniquely determine either G_{i+1} or G_{i-1}; and Π_i, although it puts strong limitations on both Π_{i+1} and Π_{i-1}, does not (usually) uniquely determine either of these partitions. The maximum number of successors to Π_i in a deletion sequence is $Ft(\Pi_i)$; but no property of Π_i permits any maximum to be set on the number of its predecessors in the sequence.

It is, however possible to devise a sequence of graphs similar in some respects to a deletion sequence, but possessing the following useful properties:

(a) Π_i uniquely determines both Π_{i+1} and Π_{i-1}, and hence every Π in the sequence.

(b) Both the maximum number of successors and the maximum number of predecessors of Π_i in the sequence are uniquely determined by Π_i.

For the construction of such a sequence, consider a connective transfusion between the connected loopless graph G_i and the unique graph $(c_y, 1, c_z)$ (for the partition 1, 1) to form the graph G_{i+1}. This particular kind of transfusion, where the edges expunged are any one cyclic edge in G_i and the only edge in $(c_y, 1, c_z)$, will be called a *one stage expansion* of G_i. Here evidently $Ft(G_{i+1}) = Ft(G_i) - 1$, and so, in an *expansion sequence* of connected loopless graphs where G_{i+1} is formed by the one stage expansion of G_i, the maximum number of successors or *expansion products* of G_i is $Ft(G_i)$. In such a sequence, G_i does not (usually) uniquely determine either G_{i+1} or G_{i-1}; but Π_i does uniquely determine both Π_{i+1} and Π_{i-1}, for Π_{i+1} is merely Π_i with two unit integers adjoined, and Π_{i-1} is merely Π_i with two unit integers omitted.

In the process of *one-stage contraction* (which is the inverse of one-stage expansion), two end points attached to two distinct points c_j and c_k of G_i (together with their end-edges) are expunged, and a new edge attaching c_j to c_k is inserted to form G_{i+1}. No sequence of one-stage contractions can be continued indefinitely, for one obvious limitation on contraction is the following. Let $Fu(\Pi)$ be the number of unit integers in Π; and let $Fu'(\Pi)$ be one half of the largest even number $\leqq Fu(\Pi)$. An algebraic expression for $Fu'(\Pi)$ is

$$Fu'(\Pi) = \tfrac{1}{4}\{2Fu(\Pi) - 1 + (1)^{Fu(\Pi)}\}.$$

The maximum number of successors of Π_i in a *contraction sequence* cannot exceed $Fu'(\Pi_i)$. But there is another limitation on the number of such succesors. The integer p_1 in Π_i is identical with the integer p_1 in Π_{i+1}; but $\Sigma\Pi_i/2 = \Sigma\Pi_{i+1}/2 + 1$. Hence $Fr(\Pi_{i+1}) = Fr(\Pi_i) - 1$. It may be that, after $Fr(\Pi_i)$ successive contractions, of which the first is performed on Π_i, the stage is reached where $Fr(\Pi_{i+Fr(\Pi_i)}) = 0$. In this case $G_{i+Fr(\Pi_i)}$ is a rosette; consequently, even if $Fu'(\Pi_{i+Fr(\Pi_i)}) > 0$, no further contraction is possible without the formation of a forbidden loop. Since the successors of Π_i in a contraction sequence are respectively the predecessors of the same Π_i in an expansion sequence, the maximum number of such predecessors is the lesser of the two functions $Fr(\Pi_i)$ and $Fu'(\Pi_i)$. (Where $Fr(\Pi_i) = Fu'(\Pi_i)$, the maximum number of predecessors is given by either of these two functions. Here the first graph in the expansion sequence is a rosette with < 2 end points.) If the lesser of $Fr(\Pi)$ and $Fu'(\Pi)$ is called $Fm(\Pi)$, then an algebraic expression for that function is

$$Fm(\Pi) = \tfrac{1}{2}\{Fr(\Pi) + Fu'(\Pi) - |(Fr(\Pi) - Fu'(\Pi)|\}.$$

A Π for which $Fm(\Pi) = 0$ will be called an *initial* Π and will be characterized by the symbol Π_0. A Π_0 is easily recognized, since either $Fr(\Pi_0) = 0$ or $Fu(\Pi_0) < 2$. There are of course Π_0's which meet both of these conditions, and so a connected loopless graph for a Π_0 is a graph which is a rosette or a graph containing < 2 end-points or both.[11]

A Π for which $Ft(\Pi) = 0$ will be called *final.* It is easy to recognize such a final Π by its defining condition. The connected loopless graphs for final Π's are trees, and a connected loopless graph for a unigraphic final Π is a unique tree. Necessary and sufficient conditions for uniqueness among trees (over and above the general condition $Ft(\Pi) = 0$ which holds for all trees) are (a) $p_1 = p_4 = 2$; or (b) $p_1 = p_3 > p_4 = 1$; or (c) $p_2 > p_3 = 1$; or (d) $p_2 = 1$. These conditions are included in the conditions (2.1), $\cdots$, (2.6″).

Let any Π which is neither initial nor final be called *intermediate.* In an intermediate Π,

$$\text{(a)}\quad Ft(\Pi) > 0 < Fr(\Pi)\,; \qquad \text{(b)}\quad p_{n-1} = 1.$$

An expansion sequence which begins with an initial and ends with a final Π will be called a *complete expansion sequence.* The number of members of a complete expansion sequence is $Fm(\Pi_i) + Ft(\Pi_i) + 1$ where Π_i is any member of the sequence. Where Π_i is a Π_0, the number of members of the complete sequence is $Ft(\Pi_0) + 1 = \mu(G_0) + 1$.

Expansion sequences lead to a very useful scheme for the classification of all Π's. Construct the disjoint classes (which together exhaust all Π's) in such a way that each class contains only complete expansion sequences. By means of this principle, the problem of classifying all Π's is reduced to the problem of classifying all initial (or all final) Π's—an incomparably easier task.

Inspection of Classes I-VII (pp. 678-679) shows that:

The unigraphic Π's of Class I are initial (sometimes also final).

The unigraphic Π's of Classes II, IVa, V and VI are initial.

[11] In constructing the successive graphs of a contraction sequence, some difficulty may be caused by the fact that occasionally, although Π_i is not initial, all of the end-points in G_i are attached to one point c_j. If this situation occurs, proceed as follows. Replace two of the end-points and their end-edges in G_i by a loop-edge dependent from c_j; then remove the loop-edge by a loop-transfusion. Such a transfusion must always be possible wherever G_i (as in the instance under consideration) is not a rosette.

The unigraphic Π's of Class III are intermediate or final.

The unigraphic Π's of Classes IVb and VII are final.

The same inspection reveals that every final Π (classes I, III, IVb or VII) belongs to a complete expansion sequence in which Π_0 is unigraphic and of Class I (where $p_2 = 1$, or $p_2 = p_1 > 1$, or $p_3 = 1$) or Class II (where $p_1 = p_3$) or Class IVa or Class VI. A Π_0 which meets any one of these conditions will be called a $\underline{\Pi}_0$.

Since necessary and sufficient conditions for uniqueness among trees are almost self-evident and have long been known, it is safe to assume that there are no unigraphic final Π's besides those in the classes listed above. Hence the

LEMMA. *A necessary and sufficient condition for a unigraphic final* Π *is that this final* Π *belong to a complete expansion sequence in which* Π_0 *is a* $\underline{\Pi}_0$.

Preservand subgraphs. Pick from any connected loopless graph G a connected subgraph V which is not the whole of G. One of two alternatives must be true:

$$\text{(a)} \quad Ft(G) = Ft(V) \quad \text{or} \qquad \text{(b)} \quad Ft(G) > Ft(V).$$

In the latter case, G may be converted by successive expansions into a connected graph G_i, where $Ft(G_i) = Ft(V)$, and G_i contains V intact. The method of expanding G to G_i is to expunge from G and its successive expansion products only cyclic edges which are not in V. The subgraph V thus preserved intact in the expansion sequence (as far as the stage where $Ft(G_i) = Ft(V)$) will be called a *preservand subgraph* of G.

Application of the idea of a preservand subgraph leads at once to the following

LEMMA. *If*

(1) *G_i and G_i' are distinct representatives of one and the same* Π_i *where* $Ft(\Pi_i) > 0$, *and*

(2) *G_i contains a preservand subgraph V which distinguishes G_i from G_i' and*

(3) *V contains no point which is an end-point in G_i, and*

(4) $Ft(V) < 2$,

then no explanation product of Π_i *is unigraphic unless* Π_i *belongs to a complete expansion sequence beginning with a* $\underline{\Pi}_0$; *and if* Π_i *does belong to a complete expansion sequence beginning with a* $\underline{\Pi}_0$, *then its only unigraphic expansion product is the final* Π.

A preservand subgraph V which meets the conditions just stated will be called a $\bar{V}$. Naturally, the above lemma is particularly useful where Π_i is a Π_0. The presence of a $\bar{V}$ in one of two distinct representatives (G_0 and G_0') of Π_0 implies that no expansion product of Π_0 is unigraphic.

Additional terminology.

(1) *Polarity.* Where $b(i,j) > 0$, if $p_i = p_j$, the bond (i,j) as well as each of the edges which together form that bond are *homopolar*; where $p_i \neq p_j$, the bond (i,j) and the edges which compose it are *heteropolar.*

(2) *Alternating and uniform partitions.* There are certain classes of partitions which, so far as the construction of one representative graph is concerned, lend themselves to particularly simple treatment. These are the *alternating* and *uniform* partitions hereafter described and denoted by the respective symbols *Pa* and *Pu.* Almost all alternating and uniform partitions are multigraphic, but each one of them has among its representative graphs one uniquely definable graph hereafter called respectively the *alternating* or the *uniform graph.* These are denoted by the respective symbols *Ga* and *Gu.* Alternating and uniform graphs are partly characterized by the facts that, in them, each point and each edge forms part of at least one cycle of order n (hereafter called a *major cycle*), and that, in them, there occurs no cycle of order k where $2 < k < n$.

(a) Alternating partitions and their alternating graphs.

A *Pa* is defined as follows:

(1) A *Pa* contains no unit integers.

(2) Any integer in *Pa* occurs there an even number of times.

Hence every *Pa* is a Π_0.

A *Ga* is defined as follows:

(1) If $n = 2$, then $b(1,2) = p_1$.

(2) If $n > 2$, then in *Ga*

$$b(i, i+1) = p_i - 1, \qquad i = 1, 3, 5, \cdots, (n-1),$$
$$b(j, j+1) = 1, \qquad j = 2, 4, 6, \cdots, (n-2), \qquad b(n, 1) = 1.$$

From these specifications, it follows that every heteropolar bond (if any) in *Ga* is single.

(b) Uniform partions and their uniform graphs.

A *Pu* is defined as follows:

(1) $p_1 = p_n$; (2) p_1 is even; n is odd.

A Gu is defined as follows:

(1) If $n = 1$, then Pu is not a Πu and Gu is the loop-rosette $(c_1, p_1/2, c_1)$.

(2) If $n > 1$, then $b(i, i+1) = p_1/2$, $i = 1, 2, \cdots, (n-1)$; $b(n, 1) = p_1/2$.

Here Pu is a Π_0 and Gu contains no heteropolar bond; moreover Gu contains no single bond unless $p_1 = 2$; in this latter case, all the bonds are single.

GENERAL RUBRIC FOR THE CLASSIFICATION OF ALL INITIAL Π'S.

	Class
$Fr(\Pi_0) = 0$	A
$Fr(\Pi_0) > 0$	
$p_1 = 2$	B
$p_1 > 2$	
$n = 3$	C
$n > 3$	
$p_1 = p_n$	D
$p_1 > p_2 = p_n$	E
$p_1 = p_{n-1} > p_n$	F
$p_1 = p_d > p_{d+1} = p_n$; $p_1 = p_2$; $p_{n-1} = p_n$	G
$p_1 > p_d > p_n$; $\Pi_0 = \Pi a + Pw$ [12]	
$\Sigma Pw = 0$	H
$\Sigma \Pi a = 0$	J
$\Sigma \Pi a > 0 < \Sigma Pw$	
$Fr(Pw) \gtreqless 0$	K
$Fr(Pw) < 0$	L

Most of the classes of initial Π's given in the above rubric require still further division into subclasses before they can be successfully treated. In fact, the most feasible scheme yet found requires 30 such subclasses. Considerations of space make it out of the question to give the necessity proof for each of these. Nor is it essential to give the necessity proof for Theorem (2) in full. Six of the simpler proofs for individual subclasses are given hereafter. The methods used in the proofs for the remaining 24 subclasses

[12] The sum of m partitions is defined as follows: If all the integers of P are divided into m disjoint subsets which respectively constitute the partitions $P_1, \cdots, P_m$, then $P = P_1 + P_2 + \cdots + P_m$. A Pw is defined as a partition in which no two integers are alike.

are similar to those here shown (for subclasses A_3, C_1, D, E_1, G_1 and J_1) though in certain instances slightly more intricate.[13]

It should be remembered that, in the six following necessity proofs for individual classes of Π's, the specifications given for each class apply *only* to the initial Π's of that class. The phrase "expand with respect to (i, j)" which is used in some of the constructions means that, in the indicated expansion, the cyclic edge to be expunged is one in the bond (i, j).

Subclass A_3. $Fr(\Pi_0) = 0, p_3 > 1$.

Π_0 is of class I, and G_0 is therefore the unique graph given (p. 678) for a Π of that class. To construct G_1, expand G_0 with respect to $(1, 2)$. To construct G_1', perform a neutral reversion on G_0 to form a graph in which $b(2,3) = 1$ and $b(1,1) = 1$; then expand this graph with respect to the unique loop-edge to form G_1'. G_1' contains an edge $(2, 3)$ terminated by two polyvalent points neither of which has the valence p_1. The tree $(c_2, 1, c_3)$ in G_1' is a $\bar{V}$. Thus the only unigraphic Π's in subclass A_3 are the initial ones.

Subclass C_1. In $\Pi_0, p_1 = p_3 > 2, n = 3$.

Π_0 is of class II, and G_0 is therefore the unique graph given (p. 678) for a Π of that class. Furthermore, every final Π in subclass C_1 is of class VII and hence has the unique graph given (p. —) for a Π of that class. In the unique G_0, $b(1,2) = b(2,3) = b(3,1) = p_1/2$. To construct G_1, expand G_0 with respect to $(2, 3)$. To construct G_1', perform a neutral reversion on G_0 to form a graph in which $b(1,1) = 1$; then expand this graph with respect to the unique loop-edge. G_1' is distinct from G_1 because, in G_1', the point c_1 is attached to two end-points, whereas in G_1 no point is attached to two end-points. Expand both G_1 and G_1' by $p_1/2 - 1$ successive expansions with respect to $(2, 3)$. Throughout this expansion, each G_i is distinct from the corresponding G_i' because G_i has a polyvalent point which is not attached to any end-point, whereas G_i' has no such polyvalent point. Any cycle of order 3 in $G'_{p_1/2}$ is a $\bar{V}$. Hence the only unigraphic Π's in subclass C_1 are the initial and final ones.

Class D. In $\Pi_0, p_1 = p_n > 2, n > 3$.

G_0 is either the alternating or the uniform graph for Π_0. To obtain G_0', construct for the partition $p_1, \cdots, p_4$ the component H_1 in which each

[13] Any reader interested in the full necessary proof for all 30 subclasses may obtain a copy of that proof by application to the author.

point terminates two single bonds and one bond of breadth $p - 2$. If $n = 4$, H_1 is G_0'. If $n > 4$, the component H_2 is either the alternating or the uniform graph for the partition $p_5, \cdots, p_n$. If $n = 5$, transfuse each loop-edge in H_2 with an edge in a multiple bond in H_1 to form G_0'. If $n > 5$, transfuse one edge in H_2 with an edge in H_1 to form G_0'. In any case, any cycle of order 3 in G_0' is a $\bar{V}$. Hence there are no unigraphic Π's in class D.

Subclass E_1. $Fr(\Pi_0) = 1, n > 3, p_1 > p_2 = p_n$.

Π_0 is of class V, and G_0 is therefore the unique graph given (p. 679) for a Π of that class. G_0 is defined as follows: $b(1,2) = b(1,3) = p_2 - 1$, $b(2,3) = 1$, $b(1,i) = p_2$, $i = 4, \cdots, n$. To obtain G_1, expand G_0 with respect to $(1,2)$; to obtain G_1', expand G_0 with respect to $(2,3)$. In G_1, the tree $(c_2, 1, c_3)$ containing a homopolar edge is a $\bar{V}$. Hence the only unigraphic Π's in subclass E_1 are the initial ones.

Subclass G_1. In $\Pi_0, p_1 = p_d > p_{d+1} = p_n, p_1 = p_2, p_{n-1} = p_n = 2$.

The component I_1 is either the alternating or the uniform graph for the partition $p_1, \cdots, p_d$. The component I_2 is the loop-rosette $(c_n, 1, c_n)$. To construct the component H_1, transfuse the loop-edge in I_2 with an edge in $(1,2)$ in I_1. The component H_2 is either the alternating or the uniform graph for the partition $p_{d+1}, \cdots, p_{n-1}$. To construct G_0, transfuse an edge in H_2 with the single edge $(1,n)$ in H_1. To construct G_0', transfuse an edge in H_2 with an edge in $(1,2)$ in H_1. Any cycle of order n in G_0 contains only two heteropolar bonds and is therefore a $\bar{V}$. Hence there are no unigraphic Π's in subclass G_1.

Subclass J_1. $Fr(\Pi_0) > 0, p_g > p_{g+1}, g = 1, \cdots, (n-1), n > 3;$

$$p_1 = (p_2 + p_3 + \cdots + p_n) - 2z, 0 < 2z < p_n.$$

Attach each point c_g $(g = 2, \cdots, (n-1))$ to c_1 by p_g edges. Attach c_n to c_1 by $p_n - 2z$ edges and complete a (not loopless) graph for Π_0 by making c_n the loop-point for z dependent loop-edges. To construct G_0, transfuse each loop-edge dependent from c_n with an edge in $(1,3)$. To construct G_0', transfuse each loop-edge dependent from c_n with an edge in $(1,2)$. Any tree $(c_3, 1, c_n)$ in G_0 is a $\bar{V}$. Hence there are no ungraphic Π's in subclass J_1.

The full necessity proof (extracts from which have just been given) completes the proof for Theorem (2).

From Theorem (2) it follows that there are just four kinds of complete expansion sequence:

(1) Those in which no Π is unigraphic.

(2) Those in which just one, the initial Π is unigraphic.

(3) Those in which just two (the initial and the final) Π's are unigraphic.

(4) Those in which more than two (and here all) of the Π's are unigraphic.

LEMMA. *If any initial Π is multigraphic, then each of its expansion-products is multigraphic. If any intermediate Π is unigraphic, then each of its expansion and contraction products is unigraphic.*

Part IV. Polychromatic graphs.

Polychromatic graphs have been defined and the conditions for the identity of two such graphs have been given above. The existence condition (1.4) for connected loopless graphs holds as well where the graph is polychromatic as where it is monochromatic. But for a graph with m colors $(m > 1)$, this condition must be strengthened by one obvious further condition.

CONDITION (1.5). $p_i = p_{i+m-1}$ for at least one p_i.

Where $m = 1$, condition (1.5) reduces to the identity $p_i = p_i$. A Π which meets condition (1.5) will be called a Π_m.

For every Π_m, there may be constructed at least one connected loopless graph involving each number of colors from 1 to m inclusive. And if, for Π_m $(m > 1)$, there are x distinct connected loopless monochromatic graphs, the number of distinct connected loopless polychromatic graphs involving y colors $(2 \lesseqgtr y \lesseqgtr m)$ must be $\gtreqless x$. Hence, if Π has no unique monochromatic connected loopless graph, it can have no unique polychromatic connected loopless graph. To obtain uniqueness conditions for connected loopless polychromatic graphs it is therefore sufficient to determine how far the uniqueness conditions for connected loopless monochromatic graphs continue to hold in the polychromatic case. The various classes of Π's which are unigraphic in the monochromatic case (cf. above) are therefore next examined.

CLASS I. The Π's in this class have rosette graphs. No matter what the number of classes of like-valent blind points in a rosette, and no matter how the colors are distributed among these classes, such a graph must always be unique.

CLASS II. A graph for Π in this class must always be unique, no matter how the colors are distributed among the like-valent points, if any such are present. Trichromism is possible where $p_1 = p_2 = p_3$; dichromism is possible where $p_1 = p_2 > p_3$ or $p_1 > p_2 = p_3$.

CLASS III. A polychromatic graph for a Π in this class cannot be unique unless $p_1 = p_2$. Where this extra condition is met, two kinds of polychromism are possible: (a) The two polyvalent points are dichromatic. Here a polychromatic graph is unique if and only if all the univalent points are monochromatic. (b) The two polyvalent points are monochromatic. Here a polychromatic graph is unique if and only if one of the univalent points is of one color and all the other univalent points are of one different color.

CLASS IVa. A polychromatic graph for a Π of this class is unique if and only if one point is of one color and all the other points are of one different color.

CLASS IVb. In a graph for a Π of this class, polychromism among the bivalent points would destroy the uniqueness of the graph, but the two end points may be of different colors without destroying this uniqueness.

CLASS V. In a graph for a Π of this class, polychromism is possible only among the points of valence p_2. But such polychromism always destroys the uniqueness of the graph, because the colors may be distributed in various ways with respect to the unique homopolar edge.

CLASS VI. In a graph for a Π of this class, polychromism is possible only among the points of valence p_1. But such polychromism always destroys the uniqueness of the graph because the colors may be distributed in various ways with respect to the unique end-point.

CLASS VII. In a graph for a Π of this class, polychromism is possible both among the univalent and among the polyvalent points. But in either case, polychromism destroys the uniqueness of the graph.

The facts just stated may be summed up in the following lemmas.

LEMMA (3.1). *All polychromatic graphs of classes* I *and* II *are unique.*

LEMMA (3.2). *No polychromatic graph of any one of the classes* V, VI *or* VII *is unique.*

LEMMA (3.3). *A polychromatic graph of class* III *or class* IV *is unique only under the special conditions just described; such a unique graph can involve no more than two different colors.*

Lemmas (3.1)-(3.3) taken together imply theorem (3) which is thus proved.

GEORGE HERBERT JONES LABORATORY, UNIVERSITY OF CHICAGO.

ON SYSTEMS OF LINEAR DIFFERENTIAL EQUATIONS.[1]

By William G. Leavitt.

1. A normal form. A system of n linear homogeneous differential equations linear in a parameter λ may be written in vector form

$$U' = (\lambda A + P)U, \tag{1}$$

with U a column vector and A and P n-square matrices. The transformation $U = TU_1$ by a unimodular matrix T is easily seen to result in an equation in U_1 of form (1), in which the coefficient of λ is $T^{-1}AT$. It is known [1] that if the elements of A and its characteristic roots are holomorphic in a closed bounded region R, then there exists a matrix T unimodular in R such that $T^{-1}AT$ is in triangular form.

It may be remarked that the theorem is also true for the case R is unbounded.[2] It is easily verfied that the method of [1] may be extended to this case without significant alteration. The only possible difficulty is the requirement [1; p. 470] of a function whose finite expansion[3] is specified at a set of points which in this case may not be finite (although without a finite accumulation point). However, the existence of such a function is established by a theorem of Mittag-Leffler [2; pp. 5-6]. Another proof which may be extended to the unbounded case is that given in [3] for general principal ideal rings. It is known that the ring of all functions holomorphic in an unbounded region satisfies all the postulates of a principal ideal ring except the infinite chain condition [Cf. 4; p. 351]. However, the proof of [3] makes no use of the chain condition and so may be applied immediately.

It will accordingly be assumed that the coefficients of the original equation satisfy the above conditions. In the following it will be supposed that the transformation has already been performed, so that in (1) A is triangular.

2. A canonical form for second order equations. Consideration will be limited henceforth to the case $n = 2$. Since A is triangular, a further transformation $U = \exp[\frac{1}{2}\lambda \int (a_{11} + a_{22})dz]U_2$, with a_{11} and a_{22} the diagonal

[1] Received July 26, 1950. Presented (in part) to the American Mathematical Society, April 30, 1949.

[2] A function is said to be holomorphic in an unbounded region R if it is holomorphic in and on the boundary of every finite portion of R.

[3] The expansion to a finite number of terms.

elements of A, results in an equation in U_2 with λ-coefficient a matrix whose diagonal elements are the negative of each other. Again assuming the equation already transformed, A may be written

$$A = \begin{pmatrix} \mu & a_{12} \\ 0 & -\mu \end{pmatrix}.$$

Consider the equation $AT = TB$, where T is unimodular and

$$B = \begin{pmatrix} \mu & \psi \\ 0 & -\mu \end{pmatrix}. \tag{2}$$

Clearly T must be triangular (except in the trivial case $A = B = 0$) and, from the unimodularity condition, its diagonal elements t_{11} and t_{22} must be units. It then follows that

$$\psi = (a_{12}t_{22} + 2\mu t_{12})/t_{11}, \tag{3}$$

with t_{12} arbitrary. Thus ψ is any associate of any member of the modular class of $a_{12} \bmod \mu$.

Now it is known [4; p. 351] that there is an integral function γ such that $\gamma a_{12} \equiv d \bmod \mu$, where d is the G. C. D. of a_{12} and μ. It is easily seen that an exponential function can be constructed having, at a given point, any specified finite expansion with non-zero constant term, by properly choosing the corresponding finite expansion of its exponent. Since γ is prime to μ, it is accordingly evident (using the theorem of Mittag-Leffler cited above) that an exponential function exists whose expansion at each zero of μ agrees with that of γ to the order of the zero [See also 2; p. 56]. Thus a unit function exists which is congruent to $\gamma \bmod \mu$ and which is therefore usable in place of γ. It follows that the elements of T can always be chosen in (3) to make $\psi = d$. In case d is an associate of μ (that is, μ a divisor of a_{12}) ψ may from (3) be chosen equal to zero. Thus in all cases a form exists with ψ either zero or a proper divisor of μ. Conversely, if a_{12} is a proper divisor of μ, then from (3) if ψ is also a divisor of μ, it must be an associate of a_{12}. Also if $a_{12} = 0$, then ψ cannot be a proper divisor of μ, and so $\psi = 0$. Thus in either case ψ is unique except for associates. The transformed equation whose λ-coefficient is the matrix of (2), with ψ either zero or a proper divisor of μ, will be termed the *canonical form* of the equation.

3. Asymptotic solutions. The expressions to which solutions of equations of form (1) are asymptotic for large values of $|\lambda|$ are known only when the characteristic roots of A are unequal throughout R.[4] For second

[4] See [5]. In applying the method to the complex case certain additional assumptions on R may be needed to ensure convergence.

order equations, in the above notation, this would imply that μ is a unit function. It is to be expected that even for second order equations the problem of finding asymptotic solutions in general will be quite complicated. In this section is to be considered the simplest case, in which it is assumed that R is bounded and that μ has a single zero of the first order (which can be taken as the origin without loss of generality). It will be assumed that the equation has been reduced to canonical form, and so, according to the results of the last section, ψ has one of the two forms 0 or 1.

Case I. $\psi = 1$. The equation to be solved has form (1) with

$$A = \begin{pmatrix} \mu & 1 \\ 0 & -\mu \end{pmatrix}.$$

Since it is assumed that R is bounded, the further transformation,

$$U = (\lambda + p_{12})^{\frac{1}{2}} \exp[\tfrac{1}{2} \int_0^z (p_{11} + p_{22})\, dz] Y,$$

is non-singular for $|\lambda|$ sufficiently large. Equation (1) then becomes

$$Y' = (\lambda A + Q) Y \tag{4}$$

where

$$Q = \begin{pmatrix} p - r & p_{12} \\ p_{21} & -p - r \end{pmatrix}, \text{ with } \begin{cases} p = \frac{1}{2}(p_{11} - p_{22}), \\ r = p'_{12}/2(\lambda + p_{12}). \end{cases}$$

The elimination of y_2 from the component equations of (4) is found to lead to an equation of form

$$y_1'' = [\lambda^2 \mu^2 + \lambda \chi_1(z) + \chi_2(z, \lambda)] y_1, \tag{5}$$

where χ_1 and χ_2 are holomorphic in R and χ_2 is bounded for $|\lambda|$ sufficiently large.

This is an equation of the type treated by Langer [6]. Thus with the added postulates [6; (i) p. 93 (second part) and (iv) p. 100] the expressions to which the solutions of (5) are asymptotic for large $|\lambda|$ are obtainable. The first component equation of (4) then furnishes a solution for y_2 and, by reversing the transformation leading to (4), asymptotic solutions are obtained for the original equation.

Case II. $\psi = 0$. Transform (1) by

$$U = \begin{bmatrix} \exp \int_0^z (\lambda\mu + p_{11})\, dz & 0 \\ 0 & \exp \int_0^z (-\lambda\mu + p_{22})\, dz \end{bmatrix} Y.$$

Equation (1) then becomes

$$(6)\qquad Y' = \begin{bmatrix} 0 & p_1 \exp(-\xi) \\ & \\ p_2 \exp(\xi) & 0 \end{bmatrix} Y, \text{ where } \begin{cases} \xi = 2\lambda \int_0^z \mu dz \\ p_1 = p_{12} \exp \int_0^z (p_{22} - p_{11}) dz \\ p_2 = p_{21} \exp \int_0^z (p_{11} - p_{22}) dz. \end{cases}$$

For some constant α, define recursively

$$(7)\qquad \begin{aligned} & y_{11}^{(i)} = \int_\alpha^z p_1 \exp(-\xi) y_{21}^{(i)} dz && (i = 0, 1, \cdots) \\ & y_{21}^{(0)} = 1, \quad y_{21}^{(i)} = \int_\alpha^z p_2 \exp(\xi) y_{11}^{(i-1)} dz && (i = 1, 2, \cdots), \end{aligned}$$

then the expressions

$$(8)\qquad y_{11} = \sum_0^\infty y_{11}^{(i)}, \qquad y_{21} = \sum_0^\infty y_{21}^{(i)},$$

are the components of a formal solution of (6).

Certain conditions are now to be assumed to ensure convergence:

(i) For some K and $\epsilon > 0$ there exists a curve joining the origin with every z for which $|z| < \epsilon$, such that $\int_0^z |dz| < K|z|$. Call the set of all such curves the class c.

(ii) The function $\int_0^z \mu dz$ is zero in R only at the origin.

In the region R_ξ corresponding to R define (in general four) sectors S_j by $-\pi/2 + j\pi \leqq \arg \xi \leqq \pi/2 + j\pi$, where $j = 0, 1, 2, 3$. Then assume, finally,

(iii) Each sector contains a point α_j which may, for each z of S_j, be connected with the origin by an ordinary curve C_z through z on which the real part $\Re(\xi)$ is monotonic. Call the set of all such curves the class C, and let $L(z)$ be the length of C_z. It will be assumed, further, that $L(z)$ is a bounded function of z over R.

Let H be an upper bound for the absolute values of the functions p_1, p_2, μ, their derivatives, $L(z)$, and $2\int_0^z \mu dz/\mu^2$. To specify the constant α in

(7) let α be α_j or 0 according as $j = 0, 2$ or $j = 1, 3$. Then for any z_1 on the curve between α and z, $\Re(\xi_1) \geqq \Re(\xi)$.

Suppose first that $|\xi| \leqq N$. By (ii) and the fact that μ has a zero of the first order at the origin, $|\int_0^z \mu dz| > |z|^2 K_1$ for some K_1. Thus $|z| < (N/K_1|\lambda|)^{\frac{1}{2}}$, and so $|\lambda| > N/\epsilon^2 K_1$ implies $|z| < \epsilon$. In sectors S_1 and S_3 $\alpha = 0$ in (7) and so the integration may be taken over a curve of class c. It is then easy to show by induction that $|y_{11}^{(i)}| \leqq M^{2i+1}|\lambda|^{-i-\frac{1}{2}}$, $|y_{21}^{(i)}| \leqq M^{2i}|\lambda|^{-i}$, where $M = HK(N/K_1)^{\frac{1}{2}} \exp N$. Hence the series in (8) converge when λ also satisfies the condition $|\lambda| > M^2$.

If $|\xi| > N$, again for sectors S_1 and S_3, the integration is taken over a curve C_z of class C. Since this is an ordinary curve, there is a last point β on the curve for which $|\xi| = N$, and the part of C_z from 0 to β may be replaced by a curve of class c. Then from (7), integrating by parts,

$$(9) \quad y_{11}^{(0)} = \int_0^\beta p_1 \exp(-\xi)\,dz + p_1 \exp(-\xi)/(-2\lambda\mu)\Big|_\beta^z + 1/(2\lambda)\int_\beta^z (p_1/\mu)' \exp(-\xi)\,dz.$$

The first integral is of the type considered above for $|\xi| \leqq N$, and so is less in absolute value than $M/|\lambda|^{\frac{1}{2}}$. In the above, $|\lambda|$ was chosen to make this ratio less than 1. It is now to be required that $|\lambda|$ also be chosen large enough to make $M/|\lambda|^{\frac{1}{2}} < H^4/N$. Now we may write $\lambda\mu^2 = \xi\mu^2/2\int_0^z \mu dz$, and since on the curve from β to z $|\xi| > N$, it follows that $1/|\lambda\mu^2| < H/N$. Also by (iii), for any z_1 on this curve, $\Re(\xi_1) \geqq \Re(\xi)$, so that $|\exp(-\xi)| \geqq |\exp(-\xi_1)|$. It may then be verified (assuming $H > 1$) that $|y_{11}^{(0)}| < 5H^4|\exp(-\xi)|/N$. Thus $y_{11}^{(0)} = E\exp(-\xi)/N$ where E is a generic symbol for any function for which $|E| < 5H^4$.

The curve of integration for $y_{21}^{(1)}$ in (7), for sectors S_1 and S_3, may be split in similar fashion. On the curve of class c from 0 to β it was shown above that $|y_{11}^{(0)}| < M/|\lambda|^{\frac{1}{2}}$. It is then seen that $y_{21}^{(1)}| < H^8/N^2 + 5H^6/N$, so that $y_{21}^{(1)} = E^2/N$.

This process may be continued, integrating by parts for each $y_{11}^{(i)}$. Then, using the fact that $y_{21}^{(i)\prime} = p_2 \exp(\xi) y_{11}^{(i-1)}$, it may be shown by induction that $y_{11}^{(i)} = E^{2i+1}\exp(-\xi)/N^{i+1}$ and then $y_{21}^{(i)} = E^{2i}/N^i$. Accordingly with a choice first of $N > (5H^4)^2$, and then $|\lambda|$ as specified above, the series in (8) converge and represent the solution.

In the sectors S_0 and S_2 the situation is somewhat less simple. The choice must now be $\alpha = \alpha_j$, so that again $|\exp(-\xi)| \geqq |\exp(-\xi_1)|$ for

any z_1 on the curve C_z of class C joining α_j with z. If there are points on C_z for which $|\xi_1| \leqq N$, the portion of the curve joining the first and last points on the curve for which $|\xi_1| = N$ may be replaced by two curves (by way of the origin) of class c. The remainder of C_z consists of one or two segments of a curve of class C on which $|\xi_1| > N$. Thus clearly an analysis similar to that above is possible showing that for $|\lambda|$ sufficiently large the series in (8) converge. Accordingly there exist solutions, in any sector, $y_{11} \sim y_{11}^{(0)}$ and $y_{21} \sim 1$.[5]

For a second solution take

$$y_{22}^{(i)} = \int_a^z p_2 \exp(\xi) y_{12}^{(i)} dz, \qquad (i = 0, 1, \cdots)$$

$$y_{12}^{(0)} = 1, \quad y_{12}^{(i)} = \int_a^z p_1 \exp(-\xi) y_{22}^{(i-1)} dz \qquad (i = 1, 2, \cdots),$$

with $\alpha = \alpha_j$ if $j = 1, 3$ and $\alpha = 0$ if $j = 0, 2$. If the curve of integration is of class C, then $\Re(\xi_1) \leqq \Re(\xi)$ for any z_1 on the curve, so that the analysis evidently parallels that above. There thus exists a second solution $y_{12} \sim 1$ and $y_{22} \sim y_{22}^{(0)}$.

Let Y_j designate a matrix of solutions (each column a solution) having the above form in sector S_j, then

$$Y_0 \sim \begin{bmatrix} \int_{\alpha_0}^z p_1 \exp(-\xi)\, dz & 1 \\ 1 & \int_0^z p_2 \exp(\xi)\, dz \end{bmatrix}.$$

From (6), $Y_0 K$ is a matrix of solutions for any constant matrix K, and so it represents the general solution. Thus each $Y_i = Y_0 K_i$ for some constant matrix K_i. Now the origin is a point common to all sectors, and hence it may be used to evaluate the constants K_i. Thus since $|Y_0| \sim -1$, it follows that $Y_i = Y_0 Y_0^{-1}(0) Y_i(0)$. The solution for the original equation may then be obtained by reversing the transformation leading to (6).

UNIVERSITY OF NEBRASKA.

---

[5] The symbol $\sim$ indicates expressions asymptotic for large $|\lambda|$. It is easy to show, also, that if $|\exp(-\xi)|$ is sufficiently large (that is, if the point is sufficiently removed from the curve $\Re(\xi) = 0$), then for $|\xi| > N$, $y_{11}^{(0)}$ is approximated by $-p_1 \exp(-\xi)/2\lambda\mu$.

BIBLIOGRAPHY.

[1] W. G. Leavitt, " A normal form for matrices whose elements are holomorphic functions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 463-472.

[2] G. Mittag-Leffler, " Sur la représentation analytique des fonctions monogènes uniformes," *Acta Mathematica*, vol. 4 (1884), pp. 1-79.

[3] W. G. Leavitt and G. Whaples, " On matrices with elements in a principal ideal ring," *Bulletin of the American Mathematical Society*, vol. 55 (1949, pp. 117-118.

[4] O. Helmer, " Divisibility properties of integral functions," *Duke Mathematical Journal*, vol. 6 (1940), pp. 345-356.

[5] G. D. Birkhoff and R. E. Langer, " The boundary problems and developments associated with a system of ordinary linear differential equations of the first order," *Proceedings of the American Academy of Arts and Sciences*, vol. 58 (1923), pp. 52-128.

[6] R. E. Langer, " The asymptotic solutions of certain linear ordinary differential equations of the second order," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 90-106.

ISOMORPHIC GROUPS OF LINEAR TRANSFORMATIONS, II.* [1]

By C. E. Rickart.

Introduction. In a previous paper [7] we have discussed the structure of isomorphisms of certain groups of linear transformations which reduce in the finite-dimensional case to full linear groups. In the present paper a similar investigation is carried out for groups of linear transformations which reduce in finite dimensions to unitary (including orthogonal) or symplectic groups. As in the previous case, it turns out that the group isomorphisms are essentially generated by isomorphisms of the underlying vector spaces on which the linear transformations operate.

In the finite-dimensional case, similar results for automorphisms have been announced by Dieudonné [2, 3].[2] Dieudonné assumes throughout that the coefficient domain is a field, that the index of the vector space is not zero, and, in the orthogonal and unitary cases, that the field is not of characteristic two. We assume the characteristic to be different from two throughout the discussion. However, in the unitary case, the coefficient domain can be any division ring with an involution, except that the characteristic must be different from two and the case of a field F_3 with exactly three elements is excluded. There is also no restriction here on the index of the vector spaces.

In view of Dieudonné's results, the main force of our discussion is obtained in the infinite-dimensional case. Therefore we have assumed throughout that the vector spaces have dimension at least equal to six. This avoids the special situations which occur for low dimensions and makes it possible to give a relatively uniform treatment of the whole problem. For similar reasons, we have not attempted a discussion of the symplectic case for characteristic two nor the unitary case for the field F_3.

In §1 vector spaces with a scalar product (called self-dual spaces) and self-adjoint involutions on such spaces are discussed. These spaces include

* Received July 24, 1950.

[1] Presented to the Society, April, 1950. We take this opportunity to express our appreciation to N. Jacobson for his encouragement and advice during the writing of this paper.

[2] Details of the proofs of these announced results will presumably appear in [1]. L. K. Hua [5] has also obtained the result for the finite-dimensional symplectic case when the field is not of characteristic two. (Added in proof.) Reference [1] appeared after this paper was written. The methods used there are, in general, quite different from ours.

the symplectic and unitary spaces as special cases. In §§**2** and **3** the special problems which arise in the unitary and symplectic cases respectively are discussed. The representation theorem for the group isomorphism is proved in §**4**.

1. Self-dual linear vector spaces. Let $\mathfrak{X}$ be a left linear vector space over a division ring $\mathfrak{D}$ and let $\lambda \to \lambda^*$ be a given involution in $\mathfrak{D}$. In other words, $(\lambda^*)^* = \lambda$, $(\lambda + \mu)^* = \lambda^* + \mu^*$, and $(\lambda\mu)^* = \mu^*\lambda^*$. The space $\mathfrak{X}$ is said to be *self-dual* provided there exists a *scalar product* (x, y) defined on $\mathfrak{X} \times \mathfrak{X}$ to $\mathfrak{D}$ possessing the following properties:

(i) $(\lambda x + \mu y, z) = \lambda(x, z) + \mu(y, z)$,

(ii) $(x, y) = 0$ *for all* x *implies* $y = 0$,

(iii) $(y, x) = \epsilon(x, y)^*$, *where* $\epsilon = \pm 1$ *and is independent of* x, y.

If $\mathfrak{X}^*$ is the right linear vector space over $\mathfrak{D}$ whose elements are identical with those of $\mathfrak{X}$, with addition defined as in $\mathfrak{X}$ and multiplication by scalars defined by $x\lambda = \lambda^* x$, then $\mathfrak{X}$ and $\mathfrak{X}^*$ are dual linear vector spaces relative to (x, y) as defined by Jacobson [6, p. 15]. An important consequence is that, if $x_1, \cdots, x_n$ are linearly independent elements of $\mathfrak{X}$, then there exist elements $y_1, \cdots, y_n$ in $\mathfrak{X}$ such that $(x_i, y_j) = \delta_{ij}$, where $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ii} = 1$ $(i, j = 1, 2, \cdots, n)$.

For reasons which will appear later, it will be assumed throughout our discussion that $\mathfrak{D}$ is not of characteristic two and furthermore, in case $\mathfrak{X}$ is unitary (as defined below), that $\mathfrak{D}$ is not equal to the field F_3 which contains exactly three elements.

Two vectors x, y in $\mathfrak{X}$ are said to be *orthogonal* provided $(x, y) = 0$. Two subspaces $\mathfrak{M}$, $\mathfrak{N}$ are said to be orthogonal provided $(x, y) = 0$ for all $x \in \mathfrak{M}$ and $y \in \mathfrak{N}$. In this case we write $\mathfrak{M} \perp \mathfrak{N}$. If $\mathfrak{M}$ is an arbitrary subspace of $\mathfrak{X}$, then $\mathfrak{M}^\perp$ will denote the set of all $x \in \mathfrak{X}$ such that $(x, y) = 0$ for every $y \in \mathfrak{M}$. We call $\mathfrak{M}^\perp$, which is a linear subspace of $\mathfrak{X}$, the *orthogonal complement*[3] of $\mathfrak{M}$. A subspace $\mathfrak{M}$ in $\mathfrak{X}$ is said to be *isotropic* if $\mathfrak{M} \cap \mathfrak{M}^\perp \neq (0)$ and is said to be *totally isotropic* if $\mathfrak{M} \subseteq \mathfrak{M}^\perp$. The maximum dimension which a totally isotropic subspace of $\mathfrak{X}$ can have is called the *index* of $\mathfrak{X}$ [4, p. 17]. A subspace of $\mathfrak{X}$ which is non-isotropic is itself a self-dual space relative to the scalar product (x, y) restricted to the subspace. A non-zero vector $u \in \mathfrak{X}$ such that $(u, u) = 0$ is called an *isotropic*

[3] $\mathfrak{M}^\perp$ is not in general an algebraic complement of $\mathfrak{M}$ in the usual sense. In fact, even if $\mathfrak{M} \cap \mathfrak{M}^\perp = (0)$, we need not have $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^\perp$ when $\mathfrak{X}$ is infinite-dimensional However, see Lemma 1.1 below.

vector. If u is an arbitrary isotropic vector in $\mathfrak{X}$, then there exists an isotropic vector $v \varepsilon \mathfrak{X}$ such that $(u, v) = 1$. In fact, let w be any vector in $\mathfrak{X}$ such that $(u, w) = 1$ and define $v = w - \frac{1}{2}(w, w)u$. Then $(u, v) = 1$ and $(v, v) = 0$. The vectors u, v are linearly independent and the two-dimensional subspace of $\mathfrak{X}$ spanned by u, v is non-isotropic. Thus every isotropic vector can be embedded in a non-isotropic plane.

A self-dual space in which every vector is isotropic is said to be *symplectic.* In this case it is not difficult to show that $(y, x) = -(x, y)$, for all x, y, and $\lambda^* = \lambda$, for all $\lambda \varepsilon \mathfrak{D}$, so that $\mathfrak{D}$ must be a field. It is well-known that the dimension of a symplectic space, if finite, is even.

A self-dual space which contains non-isotropic vectors is said to be *unitary.* If $\epsilon = -1$ in this case, then, by a trivial modification of the involution and scalar product, it is possible to obtain [4] $\epsilon = 1$. Therefore we shall always assume $\epsilon = 1$ in the unitary case. A non-isotropic subspace of a unitary space obviously cannot be symplectic and so accordingly must itself be unitary. Let $\mathfrak{M}$ be a finite-dimensional non-isotropic subspace of a unitary space $\mathfrak{X}$. Then there exists a basis $u_1, \cdots, u_n$ for $\mathfrak{M}$ consisting of non-isotropic orthogonal vectors; that is, $(u_i, u_j) = 0$ for $i \neq j$ and $(u_i, u_i) \neq 0$. This is proved as follows. Choose u_1 as any non-isotropic vector in $\mathfrak{M}$ and suppose $u_1, \cdots, u_k$, where $1 \leq k < n = \dim \mathfrak{M}$, already chosen in $\mathfrak{M}$ such that $(u_i, u_j) = 0$ for $i \neq j$ and $(u_i, u_i) \neq 0$. Observe that $u_1, \cdots, u_k$ must be linearly independent and span a non-isotropic subspace of $\mathfrak{M}$. Set [5] $\mathfrak{N} = \mathfrak{D}u_1 + \cdots + \mathfrak{D}u_k$ and note that $\dim(\mathfrak{M} \cap \mathfrak{N}^\perp) = n - k$. Since $k < n$ and [5a] $\mathfrak{M} = \mathfrak{N} \oplus (\mathfrak{M} \cap \mathfrak{N}^\perp)$, $\mathfrak{M} \cap \mathfrak{N}^\perp$ is non-zero and non-isotropic. Choose a non-isotropic vector u_{k+1} in $\mathfrak{M} \cap \mathfrak{N}^\perp$. Then $(u_i, u_{k+1}) = 0$ for $i = 1, \cdots, k$ and $(u_{k+1}, u_{k+1}) \neq 0$. The desired result now follows by induction. If $\mathfrak{X}$ is unitary and $\mathfrak{D}$ is a field with $\lambda^* = \lambda$ for all λ, then $\mathfrak{X}$ is said to be an *orthogonal* space.

Now return to a general self-dual space $\mathfrak{X}$ (that is, $\mathfrak{X}$ is either symplectic or unitary) and let A be a linear transformation on $\mathfrak{X}$. A second linear transformation A^* on $\mathfrak{X}$ such that $(xA, y) = (x, yA^*)$ for all x, y is called the *adjoint* of A. If A^* exists, then it is unique. In case $A^* = A$, then A is said to be *self-adjoint.* A linear transformation P such that $P^2 = P$ is called a *projection.* If $P^* = P$, then the range of P is a non-isotropic subspace of $\mathfrak{X}$. In fact, suppose $y \varepsilon (\mathfrak{X}P) \cap (\mathfrak{X}P)^\perp$. Then, for all x, we

---

[4] This fact was pointed out to us in conversation by N. Jacobson.

[5] The one-dimensional subspace of $\mathfrak{X}$ which contains a non-zero vector $u \varepsilon \mathfrak{X}$ will be denoted by $\mathfrak{D}u$.

[5a] See Lemma 1.1 below.

have $(x, y) = (x, yP) = (xP, y) = 0$. Therefore $y = 0$ and $\mathfrak{X}P$ is non-isotropic. Similarly $\mathfrak{X}(I - P)$ is non-isotropic, where I is the identity transformation. Furthermore $\mathfrak{X}(I - P) = (\mathfrak{X}P)^{\perp}$ and therefore $\mathfrak{X} = (\mathfrak{X}P) \oplus (\mathfrak{X}P)^{\perp}$. Conversely, if $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$ and P is the projection of $\mathfrak{X}$ onto $\mathfrak{M}$, then P is self-adjoint. Decompositions of the form $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$ do not hold in general for non-isotropic subspaces $\mathfrak{M}$ which are infinite-dimensional. However, if $\mathfrak{M}$ is finite-dimensional the decomposition does hold.

LEMMA 1.1. *If $\mathfrak{M}$ is a finite-dimensional non-isotropic subspace of $\mathfrak{X}$, then $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$.*

Proof. Since $\mathfrak{M}$ is non-isotropic, it is self-dual relative to (x, y) restricted to $\mathfrak{M}$. Therefore, if $u_1, \cdots, u_n$ is a basis for $\mathfrak{M}$, there exist $v_1, \cdots, v_n \, \varepsilon \, \mathfrak{M}$ such that $(u_i, v_j) = \delta_{ij}$. Observe that $v_1, \cdots, v_n$ must be linearly independent and hence constitute a basis for $\mathfrak{M}$. Now, for $x \, \varepsilon \, \mathfrak{X}$, write $x = u + v$, where $u = \sum_{k=1}^{n} (x, v_k) u_k$, $v = x - u$. Then for each v_j, $(v, v_j) = (x, v_j) - \sum_{k=1}^{n} (x, v_k)(u_k, v_j) = 0$. Therefore $v \, \varepsilon \, \mathfrak{M}^{\perp}$. Since $u \, \varepsilon \, \mathfrak{M}$, we have $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$.

In addition to the usual operations of addition and multiplication of linear transformations, we shall find it convenient to use the *circle operation*, $A \circ B = A + B - AB$. This operation is associative and has the identically zero transformation as an identity. Since $I - (A \circ B) = (I - A)(I - B)$, the mapping $A \to I - A$ takes the circle operation into ordinary multiplication. If a transformation A has an inverse relative to the circle operation, it will be denoted by A°. Thus $A^{\circ} \circ A = A \circ A^{\circ} = 0$ and $I - A^{\circ} = (I - A)^{-1}$, where $(I - A)^{-1}$ is the ordinary multiplicative inverse of $I - A$. We shall denote by $\mathfrak{U}(\mathfrak{X})$ the class of all linear transformations A on $\mathfrak{X}$ such that $A^* = A^{\circ}$. Observe that, if A° exists, then $A \, \varepsilon \, \mathfrak{U}(\mathfrak{X})$ if, and only if,

$$(1.1) \qquad (xA, yA) = (xA, y) + (x, yA)$$

for all x, y. $\mathfrak{U}(\mathfrak{X})$ is a group under the circle operation and is isomorphic, under the correspondence $A \to I - A$, to the multiplicative group of all linear transformations T such that $T^* = T^{-1}$; that is $(xT, yT) = (x, y)$ for all x, y. The elements of $\mathfrak{U}(\mathfrak{X})$ are called *symplectic*, *unitary* or *orthogonal* transformations according as $\mathfrak{X}$ is symplectic, unitary or orthogonal respectively.

If F is a finite-dimensional linear transformation on $\mathfrak{X}$ for which F^* exists, then it is not difficult to verify that F and F^* have the forms [6]

---

[6] For a proof that F has the indicated form, see [6, p. 17].

$$xF = \sum_{i=1}^{n} (x, b_i) a_i, \qquad xF^* = \epsilon \sum_{i=1}^{n} (x, a_i) b_i,$$

where n is the dimension of $\mathfrak{X}F$.

A linear transformation T is called an *involution* provided $T^\circ = T$; that is, $T \circ T = 0$ or $T^2 = 2T$. An involution T is in $\mathfrak{U}(\mathfrak{X})$ if, and only if, $T^* = T$. In this case $\frac{1}{2}T$ is a self-adjoint projection, so that $\mathfrak{M} = \mathfrak{X}T$ is a non-isotropic subspace and $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^\perp$. $\mathfrak{M}$ is called the *subspace of* T. Observe that T is uniquely determined by $\mathfrak{M}$ and that $xT = 2x$ for every $x \in \mathfrak{M}$ while $xT = 0$ for $x \in \mathfrak{M}^\perp$. If $\mathfrak{M}$ has minimum dimension (that is, $\dim \mathfrak{M} = 2$ or 1 according as $\mathfrak{X}$ is symplectic or unitary), then T is said to be *minimal.* If $\mathfrak{M}^\perp$ has minimal dimension, then T is said to be *maximal.* If T is either minimal or maximal, then it is said to be *extremal.* If $\mathfrak{M}$ is the subspace of an involution T in $\mathfrak{U}(\mathfrak{X})$, then $\mathfrak{M}^\perp$ is the subspace of $(2I) \circ T$. Hence T is minimal if, and only if, $(2I) \circ T$ is maximal. A one-dimensional involution U is in $\mathfrak{U}(\mathfrak{X})$ if, and only if, it has the form $xU = 2(x, u)(u, u)^{-1}u$. It follows from Lemma 1.1 that, for any finite-dimensional non-isotropic subspace $\mathfrak{M}$ of $\mathfrak{X}$, there exists an involution in $\mathfrak{U}(\mathfrak{X})$ with $\mathfrak{M}$ as its subspace. In fact, if P is the projection of $\mathfrak{X}$ onto $\mathfrak{M}$ given by Lemma 1.1, then $2P$ is the desired involution.

LEMMA 1.2. *Let T_1, T_2 be involutions in $\mathfrak{U}(\mathfrak{X})$ with subspaces $\mathfrak{M}_1$, $\mathfrak{M}_2$ respectively. A sufficient condition for $T_1 \circ T_2 = T_2 \circ T_1$ is that $\mathfrak{M}_1$ be contained in either $\mathfrak{M}_2$ or $\mathfrak{M}_2^\perp$. If T_1 is minimal, the condition is also necessary.*

Proof. The proof of the sufficiency is not difficult so will be omitted.[7] Therefore assume T_1 to be minimal and that T_1, T_2 commute. (Note that $T_1 \circ T_2 = T_2 \circ T_1$ is equivalent to $T_1T_2 = T_2T_1$.) Since T_1, T_2 commute, $\frac{1}{4}T_1T_2$ is a self-adjoint projection with range $\mathfrak{M}_1 \cap \mathfrak{M}_2$. Therefore $\mathfrak{M}_1 \cap \mathfrak{M}_2$ is non-isotropic. Since $\mathfrak{M}_1$ is minimal, it follows that either $\mathfrak{M}_1 \cap \mathfrak{M}_2 = \mathfrak{M}_1$ or $\mathfrak{M}_1 \cap \mathfrak{M}_2 = (0)$. A similar argument applied to the projection $\frac{1}{2}T_1(I - \frac{1}{2}T_2)$ shows that either $\mathfrak{M}_1 \cap \mathfrak{M}_2^\perp = (0)$ or $\mathfrak{M}_1 \cap \mathfrak{M}_2^\perp = \mathfrak{M}_1$. This completes the proof.

COROLLARY. *Let T_1, T_2 be extremal involutions with associated minimal subspaces $\mathfrak{M}_1$, $\mathfrak{M}_2$. If T_1, T_2 commute, then either $\mathfrak{M}_1 = \mathfrak{M}_2$ or $\mathfrak{M}_1 \perp \mathfrak{M}_2$.*

LEMMA 1.3. *Let P be a self-adjoint projection in $\mathfrak{X}$ and let Z be an involution in $\mathfrak{U}(\mathfrak{X})$ which commutes with every minimal involution $U \in \mathfrak{U}(\mathfrak{X})$ such that $UP = PU = U$. Then $PZ = \lambda P$, where $\lambda = 0$ or 2.*

[7] See [7, Lemma 2.2] for a proof.

Proof. If $\mathfrak{X}$ is symplectic, then every non-zero vector $u \varepsilon \mathfrak{X}P$ can be embedded in a minimal non-isotropic subspace $\mathfrak{M} \subseteq \mathfrak{X}P$. By Lemma 1.1, there exists a minimal involution $U \varepsilon \mathfrak{U}(\mathfrak{X})$ with $\mathfrak{M}$ as its subspace. Obviously $PU = UP = U$ so that Z commutes with U. Now, by Lemma 1.2, $\mathfrak{M}$ must be contained in either the subspace of Z or its orthogonal complement. It follows that $uZ = \lambda_u u$, where $\lambda_u = 0$ or 2. If u and v are any non-zero vectors in $\mathfrak{X}P$, the additivity of Z implies $(\lambda_{u+v} - \lambda_u)u = (\lambda_{u+v} - \lambda_v)v = 0$. Since the λ's assume only the values 0 and 2, it follows that $\lambda_u = \lambda_{u+v} = \lambda_v$. Therefort λ_u is a constant λ independent of u. This completes the proof for the symplectic case.

If $\mathfrak{X}$ is unitary, then the above proof gives $uZ = \lambda_u u$, where $\lambda_u = 0$ or 2, for every non-isotropic $u \varepsilon \mathfrak{X}P$. Hence let u be an isotropic vector in $\mathfrak{X}P$ and choose a second isotropic vector in $\mathfrak{X}P$ such that $(u, v) = 1$. Then $u + v$, $u - v$ are non-isotropic so that $(u + v)Z = \lambda_{u+v}(u + v)$ and $(u - v)Z = \lambda_{u-v}(u - v)$. We prove now that $\lambda_{u+v} = \lambda_{u-v}$. At this point it is necessary to use the hypothesis $\mathfrak{D} \neq F_3$, the problem being to obtain an $\alpha \varepsilon \mathfrak{D}$ such that $\alpha \neq \pm 1$ and $\alpha + \alpha^* \neq 0$. Since $\mathfrak{D} \neq F_3$, there exists a non-zero $\beta \varepsilon \mathfrak{D}$ such that $\beta \neq \pm 1$. If $\beta + \beta^* \neq 0$, take $\alpha = \beta$. If $\beta + \beta^* = 0$, take $\alpha = \beta + 1$. Then $\alpha + \alpha^* = 2$ and clearly $\alpha \neq \pm 1$. Thus the desired α exists. With this choice of α, the element $u + \alpha v$ is non-isotropic, so that $(u + \alpha v)Z = \lambda_{u+\alpha v}(u + \alpha v)$. On the other hand, if $\lambda_{u+v} \neq \lambda_{u-v}$, then either $uZ = u - v$ and $vZ = v - u$ or $uZ = u + v$ and $vZ = u + v$, according as $\lambda_{u+v} = 0$ or 2. Therefore $(u + \alpha v)Z = (1 - \alpha)(u - v)$ or $(u + \alpha v)Z = (1 + \alpha)(u + v)$ according as $\lambda_{u+v} = 0$ or 2. Since u, v are linearly independent, it follows that $\lambda_{u+\alpha v} = 1 - \alpha$ or $\lambda_{u+\alpha v} = 1 + \alpha$. But $\lambda_{u+\alpha v} = 0$ or 2 so that either possibility contradicts $\alpha \neq \pm 1$. It follows that $\lambda_{u+v} = \lambda_{u-v}$ and hence $(u + v)Z = \lambda_{u+v}(u + v)$ and also $(u - v)Z = \lambda_{u+v}(u - v)$. Adding these equations and dividing by two, we obtain $uZ = \lambda_{u+v}u$ or $uZ = \lambda_u u$, where $\lambda_u = 0$ or 2. Again it follows that λ_u is a constant λ independent of u and this completes the proof.

Lemma 1.3 is not true in general without the condition [8] $\mathfrak{D} \neq F_3$ and is the only place in our discussion where this condition is needed. On the other hand, an examination of the proof shows that $\mathfrak{D} = F_3$ can be admitted if there are no isotropic vectors in $\mathfrak{X}P$ and in particular if $\mathfrak{X}$ has index zero. It is also not difficult to prove the lemma when $\mathfrak{D} = F_3$ provided $\dim(\mathfrak{X}P) > 2$.

Consider now a subgroup $\mathfrak{G}$ of the group $\mathfrak{U}(\mathfrak{X})$ which contains all of the minimal involutions in $\mathfrak{U}(\mathfrak{X})$. The objective in the remainder of this

[8] See, e.g., [4, p. 24].

section is to obtain a characterization of the extremal involutions in $\mathfrak{G}$ in terms only of the group operation in $\mathfrak{G}$.

Let $\mathfrak{S}$ denote a set of involutions in $\mathfrak{G}$ and denote by $c(\mathfrak{S})$ the set of all involutions in $\mathfrak{G}$ each of which commutes with every element of $\mathfrak{S}$. Also, if T_1, T_2 are involution in $\mathfrak{G}$, denote by $\rho(T_1, T_2)$ the number of distinct involutions in $c(c(T_1, T_2))$. In addition, define

$$\rho = \max \rho(T_1, T_2), \qquad T_1 \circ T_2 = T_2 \circ T_1,$$

$$\rho_T = \max \rho(T, T'), \qquad T \circ 'T = T' \circ T.$$

THEOREM 1.4. *If $\mathfrak{G}$ contains involutions which are not extremal, then a necessary and sufficient condition for an involution $T \varepsilon \mathfrak{G}$ to be extremal is that $\rho_T = \frac{1}{2}\rho$. A necessary and sufficient condition for every involution in $\mathfrak{G}$ to be extremal is that $\rho_T = \rho$ for every T.*

Proof. By hypothesis $\dim \mathfrak{X} \geq 6$; therefore the only case in which all involutions are extremal is the symplectic case with $\dim \mathfrak{X} = 6$. Hence, if $\mathfrak{X}$ is symplectic, assume $\dim \mathfrak{X} > 6$. Let T_1, T_2 be two commutative involutions in $\mathfrak{G}$ and set $E_i = \frac{1}{2}T_i$ $(i = 1, 2)$. The E_i are self-adjoint projections and, since E_1, E_2 commute, the following are also self-adjoint projections:

$$P_1 = E_1E_2, \qquad P_2 = E_1(I - E_2),$$

$$P_3 = (I - E_1)E_2, \qquad P_4 = (I - E_1)(I - E_2).$$

Evidently $P_iP_j = 0$ for $i \neq j$ and $I = \Sigma P_i$. Observe that $c(T_1, T_2)$ consists of all involutions $T \varepsilon \mathfrak{G}$ such that T commutes with each P_i. Now let U_i be any minimal involution in $\mathfrak{G}$ such that $P_iU_i = U_iP_i = U_i$. Then $P_jU_i = U_iP_j = 0$, for $i \neq j$, so that $U_i \varepsilon c(T_1, T_2)$. It follows from Lemma 1.3 that every $T \varepsilon c(c(T_1, T_2))$ is of the form $T = \Sigma \delta_i P_i$, where $\delta_i = 0$ or 2. Conversely, every T of this form is in $c(c(T_1, T_2))$ provided only that $T \varepsilon \mathfrak{G}$. If none of the P_i is zero, it follows by direct calculation that $\rho(T_1, T_2) \leq 16$ or 8 according as $2I \varepsilon \mathfrak{G}$ or $2I \notin \mathfrak{G}$. On the other hand, if one of the P_i is zero, then $\rho(T_1, T_2) \leq 8$ or 4 according as $2I \varepsilon \mathfrak{G}$ or $2I \notin \mathfrak{G}$.

Next let T_1 be an arbitrary involution in $\mathfrak{G}$ with subspace $\mathfrak{M}$, where $\mathfrak{M} \neq (0)$ and $\mathfrak{M}^\perp \neq (0)$. Let $\mathfrak{M}_1$ be a minimal non-isotropic subspace of $\mathfrak{M}$ and $\mathfrak{M}_2$ a minimal non-isotropic subspace of $\mathfrak{M}^\perp$. Also let U_i be a minimal involution with subspace $\mathfrak{M}_i$ $(i = 1, 2)$ and set $T_2 = U_1 \circ U_2$. Evidently U_1, U_2 commute with one another and also with T_1. Therefore T_2 is an involution in $\mathfrak{G}$ which commutes with T_1. With T_2 chosen in this way, at most one of the associated projections P_i is zero and this happens if, and

only if, T_1 is extremal. It follows that $\rho(T_1, T_2) = \rho$ if, and only if, T_1 is not extremal and $\rho(T_1, T_2) = \frac{1}{2}\rho$ if, and only if, T_1 is extremal.[9]

The last statement of the theorem follows from the above arguments plus the fact that every involution in $\mathfrak{G}$ is extremal if, and only if, $\mathfrak{X}$ is symplectic of dimension six.

In the remainder of this paper we shall be concerned with two self-dual spaces $\mathfrak{X}$ and $\mathfrak{Y}$ of the same type (that is, either both symplectic or both unitary) over division rings $\mathcal{D}$ and $\mathcal{E}$ respectively. The same notations for the involutions and scalar products will be used for both spaces. It will be assumed throughout that the dimensions of $\mathfrak{X}$ and $\mathfrak{Y}$ are at least six, that neither $\mathcal{D}$ nor $\mathcal{E}$ has characteristic two and, in the unitary case, that both $\mathcal{D}$ and $\mathcal{E}$ are different from F_3.

Let $\mathfrak{G}$ and $\mathfrak{H}$ be subgroups of $\mathcal{U}(\mathfrak{X})$ and $\mathcal{U}(\mathfrak{Y})$ respectively, each of which contains all of the minimal involutions. We assume given a group isomorphism $G \rightarrow \mathfrak{g}(G)$ of $\mathfrak{G}$ onto $\mathfrak{H}$. In view of Lemma 1.3, $\mathfrak{H}$ will contain $2I$ if, and only if, $\mathfrak{G}$ contains $2I$. Furthermore, if $2I \varepsilon \mathfrak{G}$, then $\mathfrak{g}(2I) = 2I$. As a consequence of Theorem 1.4, we have that $\mathfrak{g}(G)$ is an extremal involution in $\mathfrak{H}$ if, and only if, G is an extremal involution in $\mathfrak{G}$. The purpose of the next three sections is to obtain a representation of the group isomorphism $\mathfrak{g}$. The next part of our discussion has to deal separately with the unitary and symplectic cases.

2. The unitary case. Throughout this section, both $\mathfrak{X}$ and $\mathfrak{Y}$ are assumed to be unitary. We prove first two lemmas for unitary spaces.

LEMMA 2.1.[10] *Let u, v_i be isotropic vectors in $\mathfrak{X}$ such that $(u, v_i) = 1$, for $i = 1, \cdots, k$, where $k + 2 < \dim \mathfrak{X}$. Then there exists an isotropic vector w linearly independent of $u, v_i, \cdots, v_k$ and such that $(u, w) = 1$.*

Proof. Since $\dim (\mathcal{D}w)^{\perp} = \dim \mathfrak{X} - 1 > k + 1$, there exists $r \varepsilon (\mathcal{D}u)^{\perp}$ linearly independent of $u, v_1, \cdots, v_k$. Set $w = \frac{1}{2}(s, s)u + s$, where $s = v_1 + r$. Then $(w, w) = 0$, $(u, w) = 1$ and w is linearly independent of $u, v_1, \cdots, v_k$.

LEMMA 2.2. *Let $u, v_1, \cdots, v_k$ be vectors in $\mathfrak{X}$ such that u is linearly*

[9] Observe that in this case $2P_1 = U_1$, $2P_2 = T_1 \circ U_1$, $2P_3 = U_2$, and $2P_4 = (2I) \circ T_1 \circ U_2$. Also $\Sigma\delta_i P_i = (\delta_1 P_1) \circ (\delta_2 P_2) \circ (\delta_3 P_3) \circ (\delta_4 P_4)$. Therefore, if we set

$$V = (\epsilon_1 U_1) \circ (\epsilon_2 (T_1 \circ U_2)) \circ (\epsilon_3 U_2) \text{ and } W = V \circ T_1 \circ U_2,$$

where $\epsilon_i = \frac{1}{2}\delta_i$, then $V, W \varepsilon \mathfrak{G}$ and $\Sigma\delta_i P_i = V$ or $2I \circ W$ according as $\epsilon_4 = 0$ or 1.

[10] This lemma, and the lemma which follows, are needed below for the case $k = 3$. Hence the restriction $\dim \mathfrak{X} \geqq 6$ in the unitary case.

independent of $v_1, \cdots, v_k$, *where* $0 \leq 2k \leq \dim \mathfrak{X}$ *and at least one of the vectors* $v_1, \cdots, v_k$ *is non-isotropic. Then there exists a non-isotropic vector* w *such that* $(u, w) = 1$ *and* $(v_i, w) = 0$ *for* $i = 1, \cdots, k$.

Proof. Choose $r \varepsilon \mathfrak{X}$ such that $(u, r) = 1$ and $(v_i, r) = 0$ for $i = 1, \cdots, k$. If $(r, r) \neq 0$, we can take $w = r$. Therefore assume $(r, r) = 0$. Since at least one of the vectors v_i is non-isotropic, the subspace $\mathfrak{D}v_1 + \cdots + \mathfrak{D}v_k$ is not contained in $(\mathfrak{D}v_1 + \cdots + \mathfrak{D}v_k)^{\perp}$. Hence the intersection of these subspaces has dimension at most $k - 1$. On the other hand, since $2k \leq \dim \mathfrak{X}$, we have $\dim (\mathfrak{D}v_1 + \cdots + \mathfrak{D}v_k)^{\perp} \geqq \dim \mathfrak{X} - k \geqq k$. It follows that there exists a vector s linearly independent of $v_1, \cdots, v_k$ such that $(v_i, s) = 0$ for $i = 1, \cdots, k$. Now, if r is linearly dependent on $v_1, \cdots, v_k$, consider $r' = r - (s, u)r + s$. Then r' is linearly independent of $v_1, \cdots, v_k$, $(u, r') = 1$ and $(v_i, r') = 0$ for $i = 1, \cdots, k$. Therefore we may as well assume r to be linearly independent of $v_1, \cdots, v_k$. It is easy to see that r must also be linearly independent of $u, v_1, \cdots, v_k$. Next choose $t \varepsilon \mathfrak{X}$ such that $(r, t) = 1$, $(u, t) = 0$ and $(v_i, t) = 0$ for $i = 1, \cdots, k$. Define $w = r + \lambda t$, where λ is to be determined. Note first that $(v_i, w) = 0$, for $i = 1, \cdots, k$, and $(u, w) = 1$ independently of λ. Furthermore $(w, w) = \lambda + \lambda^* + \lambda(t, t)\lambda^*$. If $(t, t) = 0$, take $\lambda = 1$ to obtain $(w, w) = 2$. If $(t, t) \neq 0$, take $\lambda = -(t, t)^{-1}$ to obtain $(w, w) = -(t, t)^{-1}$. In either case w is non-isotropic and the proof is complete.

LEMMA 2.3. *Let T be an arbitrary two-dimensional involution in $\mathfrak{U}(\mathfrak{X})$ with $\mathfrak{M}$ as its subspace. If u is any non-isotropic vector in $\mathfrak{M}$, then T can be written in the form $T = U \circ V$, where U and V are minimal involutions in $\mathfrak{U}(\mathfrak{X})$, $UV = VU = 0$ and $\mathfrak{X}U = \mathfrak{D}u$.*

Proof. There exists a non-isotropic $v \varepsilon \mathfrak{M}$ such that $(u, v) = 0$. Define $xU = 2(x, u)(u, u)^{-1}u$ and $xV = 2(x, v)(v, v)^{-1}v$. Then U, V are minimal involutions in $\mathfrak{U}(\mathfrak{X})$. Also $UV = VU = 0$ so that $U \circ V = U + V$. Since $U \circ V$ is a two-dimensional involution in $\mathfrak{U}(\mathfrak{X})$ with $\mathfrak{D}u + \mathfrak{D}v = \mathfrak{M}$ as its subspace. It follows that $T = U \circ V$.

LEMMA 2.4. *Let T be a two-dimensional involution in $\mathfrak{G}$. Then either $\mathfrak{g}(T)$ or $(2I) \circ \mathfrak{g}(T)$ is a two-dimensional involution in $\mathfrak{H}$.*

Proof. Write T in the form $T = U \circ V$, where U and V are minimal involutions. Then $\mathfrak{g}(T) = \mathfrak{g}(U) \circ \mathfrak{g}(V)$. Since $\mathfrak{g}(U)$ and $\mathfrak{g}(V)$ are extremal, it follows that either $\mathfrak{g}(T)$ or $(2I) \circ \mathfrak{g}(T)$ is two-dimensional.

Let T_1, T_2 be two-dimensional involutions in $\mathfrak{U}(\mathfrak{X})$ with subspaces $\mathfrak{M}_1$,

$\mathfrak{M}_2$ respectively. The involutions T_1, T_2 are said to *intersect* provided $T_1 \neq T_2$ and $\mathfrak{M}_1 \cap \mathfrak{M}_2 \neq (0)$. Observe that T_1, T_2 intersect if, and only if, $\dim (\mathfrak{M}_1 + \mathfrak{M}_2) = 3$.

LEMMA 2.5. *Let T_1 and T_2 be two-dimensional involutions in $\mathfrak{G}$. Then the associated two-dimensional involutions in $\mathfrak{H}$ intersect if, and only if, T_1 and T_2 intersect.*

Proof. Let $T_i = U_i \circ V_i$, where U_i, V_i are minimal involutions. It will be sufficient, because of the symmetry, to prove that intersection of T_1 and T_2 implies intersection of the associated two-dimensional involutions in $\mathfrak{H}$. If T_1 and T_2 are assumed to intersect, there is clearly no loss in assuming $\mathfrak{X}V_2$ contained in $\mathfrak{X}U_1 + \mathfrak{X}U_2 + \mathfrak{X}V_1$. Denote by $\mathfrak{N}_i$ the two-dimensional subspace associated with $\mathfrak{g}(T_i)$. Evidently $\mathfrak{N}_1 \neq \mathfrak{N}_2$; otherwise either $\mathfrak{g}(T_1) = \mathfrak{g}(T_2)$ or $\mathfrak{g}(T_1) = (2I) \circ \mathfrak{g}(T_2)$ which implies either $T_1 = T_2$ or $T_1 = (2I) \circ T_2$, neither of which is possible. Therefore we have only to prove $\dim (\mathfrak{N}_1 + \mathfrak{N}_2) = 3$. Choose vectors $r_i, s_i \varepsilon \mathfrak{Y}$ such that $\mathcal{E}r_i$ and $\mathcal{E}s_i$ are the one-dimensional subspaces of $\mathfrak{Y}$ associated with $\mathfrak{g}(U_i)$ and $\mathfrak{g}(V_i)$ respectively. Then $\mathfrak{N}_i = \mathcal{E}r_i + \mathcal{E}s_i$; hence, if $\dim (\mathfrak{N}_1 + \mathfrak{N}_2) \neq 3$, then r_1, s_1, r_2, s_2 are linearly independent non-isotropic vectors in $\mathfrak{Y}$. By Lemma 2.2, there exists a non-isotropic vector $t \varepsilon \mathfrak{Y}$ such that $(r_1, t) = (s_1, t) = (r_2, t) = 0$ while $(s_2, t) = 1$. Let W be a minimal involution in $\mathfrak{G}$ such that $\mathcal{E}t$ is the one-dimensional subspace associated with the extremal involution $\mathfrak{g}(W)$. Evidently $\mathfrak{g}(W)$ commutes with and is different from each of the involutions $\mathfrak{g}(U_1)$, $\mathfrak{g}(V_1)$, $\mathfrak{g}(U_2)$. Therefore W commutes with and is different from each of the involutions U_1, V_1, U_2. It follows by the Corollary to Lemma 1.2, that $\mathfrak{X}W$ is orthogonal to each of the subspaces $\mathfrak{X}U_1$, $\mathfrak{X}V_1$, $\mathfrak{X}U_2$. Since, by hypothesis, $\mathfrak{X}V_2$ is contained in $\mathfrak{X}U_1 + \mathfrak{X}V_1 + \mathfrak{X}U_2$, it is also true that $\mathfrak{X}W$ is orthogonal to $\mathfrak{X}V_2$. Therefore $W \neq V_2$ and W, V_2 commute. It follows that $\mathfrak{g}(W) \neq \mathfrak{g}(V_2)$ and $\mathfrak{g}(W)$, $\mathfrak{g}(V_2)$ commute. But commutativity of $\mathfrak{g}(W)$, $\mathfrak{g}(V_2)$ implies that either $\mathcal{E}t = \mathcal{E}s_2$ or $(\mathcal{E}t) \perp (\mathcal{E}s_2)$. The second possibility is ruled out since $(s_2, t) = 1$. Hence $\mathcal{E}t = \mathcal{E}s_2$. Since $\mathfrak{g}(W) \neq \mathfrak{g}(V_2)$, this implies that $\mathfrak{g}(W) = (2I) \circ \mathfrak{g}(V_2) = \mathfrak{g}((2I) \circ V_2)$. Therefore $W = (2I) \circ V_2$. But this is impossible since both W and V_2 are minimal. It follows that $\dim (\mathfrak{N}_1 + \mathfrak{N}_2) = 3$ and the proof is complete.

The group isomorphism $G \to \mathfrak{g}(G)$ will now be used to construct a one-to-one mapping of the one-dimensional subspaces of $\mathfrak{X}$ onto the one-dimensional subspaces of $\mathfrak{Y}$.

THEOREM 2.6. *There exists a one-to-one mapping F of the one-*

dimensional subspaces of $\mathfrak{X}$ onto the one-dimensional subspaces of $\mathfrak{Y}$ with the following properties:

(i) *If U is an extremal involution in $\mathfrak{G}$ with associated one-dimensional subspace $\mathfrak{D}u$, then $(\mathfrak{D}u)F$ is the one-dimensional subspace associated with $\mathfrak{g}(U)$.*

(ii) *$(\mathfrak{D}u)F \perp (\mathfrak{D}v)F$ if, and only if, $\mathfrak{D}u \perp \mathfrak{D}v$.*

(iii) *$(\mathfrak{D}u_3)F \subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$ if, and only if, $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$.*

Proof. Consider first a non-isotropic one-dimensional subspace $\mathfrak{D}u$ in $\mathfrak{X}$ and let U be the minimal involution in $\mathfrak{G}$ which has $\mathfrak{D}u$ as its subspace. Then $\mathfrak{g}(U)$ is an extremal involution and therefore either the subspace $\mathfrak{N}$ of $\mathfrak{g}(U)$ or its orthogonal complement $\mathfrak{N}^{\perp}$ is a non-isotropic one-dimensional subspace of $\mathfrak{Y}$. In the first case we define $(\mathfrak{D}u)F = \mathfrak{N}$ and in the second $(\mathfrak{D}u)F = \mathfrak{N}^{\perp}$. This evidently establishes a one-to-one correspondence between the non-isotropic one dimensional subspaces of $\mathfrak{X}$ and $\mathfrak{Y}$.

If $\mathfrak{D}u$ is isotropic, then, by Lemma 2. 1, it is possible to choose two isotropic vectors v_1, v_2 such that u, v_1, v_2 are linearly independent and $(u, v_1) = (u, v_2) = 1$. Then $\mathfrak{M}_i = \mathfrak{D}u + \mathfrak{D}v_i$ $(i = 1, 2)$ are non-isotropic, two-dimensional subspaces of $\mathfrak{X}$ which intersect in $\mathfrak{D}u$. Let T_i be the involution with $\mathfrak{M}_i$ as its subspace. Then T_1, T_2 intersect and, by Lemma 2. 5, the two-dimensional subspaces associated with $\mathfrak{g}(T_1)$ and $\mathfrak{g}(T_2)$ intersect in a one-dimensional subspace of $\mathfrak{Y}$ which we define to be $(\mathfrak{D}u)F$. Observe that $(\mathfrak{D}u)F$ is isotropic; otherwise it would be possible to find extremal involutions R, S_i in $\mathfrak{H}$ such that $\mathfrak{g}(T_i) = R \circ S_i$ and $\mathfrak{Y}R = (\mathfrak{D}u)F$. But then $T_i = U \circ V_i$, where U, V_i are minimal involutions in $\mathfrak{G}$, and this implies that $\mathfrak{D}u = \mathfrak{M}_1 \cap \mathfrak{M}_2 = \mathfrak{X}U$, contradicting the assumption that $\mathfrak{D}u$ be isotropic.

We now have $(\mathfrak{D}u)F$ defined for all $\mathfrak{D}u$. The next step is to prove that F is single-valued. This is obvious if $\mathfrak{D}u$ is non-isotropic. Therefore let $\mathfrak{D}u$ be isotropic and let T_1, T_2 be the intersecting two-dimensional involutions used in the definition of $(\mathfrak{D}u)F$. The problem is to show that $(\mathfrak{D}u)F$ is independent of the choice of T_1, T_2. This is equivalent to showing that, if T_3 is any other two-dimensional involution in $\mathfrak{G}$ for which $\mathfrak{D}u \subseteq \mathfrak{X}T_3$, then the two-dimensional subspace associated with $\mathfrak{g}(T_3)$ contains $(\mathfrak{D}u)F$. Denote the two-dimensional subspace of T_i by $\mathfrak{M}_i$ and the two-dimensional subspace associated with $\mathfrak{g}(T_i)$ by $\mathfrak{N}_i$. Consider first the case in which $\mathfrak{M}_3 \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_2$; then $\dim(\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 4$. We have, by definition, that $(\mathfrak{D}u)F \subseteq \mathfrak{N}_1 \cap \mathfrak{N}_2$ and, by Lemma 2. 5, the spaces $\mathfrak{N}_1$, $\mathfrak{N}_2$, $\mathfrak{N}_3$ intersect

by pairs. Therefore it will be sufficient to prove that $\dim(\mathfrak{N}_1 + \mathfrak{N}_2 + \mathfrak{N}_3) = 4$ since this will imply $\dim(\mathfrak{N}_1 \cap \mathfrak{N}_2 \cap \mathfrak{N}_3) = 1$, which in turn will imply $(\mathfrak{D}u)F \subseteq \mathfrak{N}_3$. Let $T_i = R_i \circ S_i$, where R_i and S_i are minimal involutions in $\mathfrak{G}$. Observe that $\mathfrak{M}_i = \mathfrak{X}R_i + \mathfrak{X}S_i$ and also that $\mathfrak{N}_i = (\mathfrak{X}R_i)F + (\mathfrak{X}S_i)F$. Since $\dim(\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 4$, we have $\dim(\mathfrak{X}R_1 + \mathfrak{X}S_1 + \mathfrak{X}R_2 + \mathfrak{X}R_3) = 4$. Hence, by Lemma 2. 2, there exists a one-dimensional non-isotropic subspace $\mathfrak{D}w$ in $\mathfrak{X}$ which is orthogonal to $\mathfrak{X}R_1$, $\mathfrak{X}S_1$ and $\mathfrak{X}R_2$ but not to $\mathfrak{X}R_3$. Then, as in the proof of Lemma 2. 5, $(\mathfrak{D}w)F$ is orthogonal to $(\mathfrak{X}R_1)F$, $(\mathfrak{X}S_1)F$ and $(\mathfrak{X}R_2)F$ but not to $(\mathfrak{X}R_3)F$. This shows that $(\mathfrak{X}R_3)F$ is not contained in $(\mathfrak{X}R_1)F + (\mathfrak{X}S_1)F + (\mathfrak{X}R_2)F$. Similarly none of the subspaces $(\mathfrak{X}R_1)F$, $(\mathfrak{X}S_1)F$, $(\mathfrak{X}R_2)F$, $(\mathfrak{X}R_3)F$ is contained in the union of the remaining three. In other words these subspaces are linearly independent. Therefore it follows that $\dim(\mathfrak{N}_1 + \mathfrak{N}_2 + \mathfrak{N}_3) = 4$ and hence $(\mathfrak{D}u)F \subseteq \mathfrak{N}_3$.

Now assume that $\mathfrak{M}_3 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Let v_i be isotropic vectors such that $(u, v_i) = 1$ and $\mathfrak{M}_i = \mathfrak{D}u + \mathfrak{D}v_i$ $(i = 1, 2, 3)$. By Lemma 2. 1, there exists an isotropic vector v_4 linearly independent of u, v_1, v_2, v_3 and such that $(u, v_4) = 1$. Set $\mathfrak{M}_4 = \mathfrak{D}u + \mathfrak{D}v_4$. Then $\mathfrak{M}_4 \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_2$ and $\mathfrak{M}_3 \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_4$. Let T_4 be the two-dimensional involution in $\mathfrak{G}$ with subspace $\mathfrak{M}_4$ and denote by $\mathfrak{N}_4$ the two-dimensional subspace associated with $\mathfrak{g}(T_4)$. Then, by the above argument, $(\mathfrak{D}u)F \subseteq \mathfrak{N}_4$ and therefore $(\mathfrak{D}u)F = \mathfrak{N}_1 \cap \mathfrak{N}_4$. Another application of the same argument, using T_4 instead of T_2 gives $(\mathfrak{D}u)F \subseteq \mathfrak{N}_3$ and completes the proof that F is single-valued. By symmetry, F is a one-to-one mapping of the one-dimensional subspaces of $\mathfrak{X}$ *onto* the one-dimensional subspaces of $\mathfrak{Y}$.

Property (i) is immediate from the definition of F. Property (ii), for non-isotropic subspaces, has already been observed in the proof of Lemma 2. 5. Next let $\mathfrak{D}u$ be isotropic, $\mathfrak{D}v$ non-isotropic, and $\mathfrak{D}u \perp \mathfrak{D}v$. Choose a vector $r \in \mathfrak{X}$ such that $(u, r) = 1$, $(v, r) = 0$ and set $w = -\frac{1}{2}(r, r)u + r$. Then $(u, w) = 1$ and $(w, w) = 0$. Evidently $u + w$, $u - w$ are non-isotropic and orthogonal to v. Therefore $(\mathfrak{D}v)F$ is orthogonal to $(\mathfrak{D}(u + w))F$ and $(\mathfrak{D}(u - w))F$. Since $\mathfrak{D}u \subseteq \mathfrak{D}(u + w) + \mathfrak{D}(u - w)$, the proof that F is single-valued gives $(\mathfrak{D}u)F \subseteq (\mathfrak{D}(u + w))F + (\mathfrak{D}(u - w))F$. Therefore $(\mathfrak{D}u)F \perp (\mathfrak{D}v)F$. Now let both $\mathfrak{D}u$ and $\mathfrak{D}v$ be isotropic with $\mathfrak{D}u \perp \mathfrak{D}v$. Choose w as before; then $\mathfrak{D}v$ is orthogonal to $\mathfrak{D}(u + w)$ and $\mathfrak{D}(u - w)$. Hence, by the result just obtained, $(\mathfrak{D}v)F$ is orthogonal to $(\mathfrak{D}(u + w))F$ and $(\mathfrak{D}(u - w))F$. Therefore again $(\mathfrak{D}u)F \perp (\mathfrak{D}v)F$. Property (ii) now follows by a symmetry argument.

In order to prove (iii), it will be sufficient to prove that $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$ implies $(\mathfrak{D}u_3)F \subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$, where $\mathfrak{D}u_3 \neq \mathfrak{D}u_1$, $\mathfrak{D}u_3 \neq \mathfrak{D}u_2$ and

hence $\mathfrak{D}u_1 \neq \mathfrak{D}u_2$. Suppose $(\mathfrak{D}u_3)F \not\subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$. Then there exists $\mathfrak{D}u_4$ such that $(\mathfrak{D}u_4)F$ is orthogonal to $(\mathfrak{D}u_1)F$ and $(\mathfrak{D}u_2)F$ but not to $(\mathfrak{D}u_3)F$. By (ii), this implies that $\mathfrak{D}u_4$ is orthogonal to $\mathfrak{D}u_1$ and $\mathfrak{D}u_2$ but not to $\mathfrak{D}u_3$. This contradicts $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$ and completes proof of the lemma.

3. The simplectic case. Throughout this section both $\mathfrak{X}$ and $\mathfrak{Y}$ are assumed to be symplectic. The objective here is to obtain the analogue of Theorem 1. 6 for this case. Recall that all of the minimal involutions in the symplectic case are two-dimensional.

LEMMA 3. 1. *Let T_1, T_2 be non-commutative minimal involutions in $\mathfrak{G}$ with subspaces $\mathfrak{M}_1, \mathfrak{M}_2$ respectively. A necessary and sufficient condition for an involution T in $\mathfrak{G}$ to belong to $c(T_1, T_2)$ is that $\mathfrak{M}_1 + \mathfrak{M}_2$ be contained in either the subspace of T or its orthogonal complement.*

Proof. The sufficiency is immediate from the first part of Lemma 1. 2. On the other hand, if $T \varepsilon c(T_1, T_2)$ and has $\mathfrak{M}$ as its subspace, then by the second part of Lemma 1. 2, either $\mathfrak{M}_i \subseteq \mathfrak{M}$ or $\mathfrak{M}_i \subseteq \mathfrak{M}^\perp$. We have only to show that the same case occurs for both $i = 1, 2$. Suppose $\mathfrak{M}_1 \subseteq \mathfrak{M}$ and $\mathfrak{M}_2 \subseteq \mathfrak{M}^\perp$. Since $\mathfrak{M}_1 \subseteq \mathfrak{M}$, we have $\mathfrak{M}^\perp \subseteq \mathfrak{M}_1{}^\perp$. Therefore $\mathfrak{M}_2 \subseteq \mathfrak{M}_1{}^\perp$ and hence $\mathfrak{M}_1 \subseteq \mathfrak{M}_2{}^\perp$. This implies $T_1T_2 = T_2T_1 = 0$ which is contrary to the hypothesis that T_1, T_2 are non-commutative. It follows that either $\mathfrak{M}_1 + \mathfrak{M}_2 \subseteq \mathfrak{M}$ or $\mathfrak{M}_1 + \mathfrak{M}_2 \subseteq \mathfrak{M}^\perp$.

LEMMA 3. 2. *Let T_1, T_2 be non-commutative extremal involutions in $\mathfrak{G}$. A sufficient condition for $T \varepsilon c(c(T_1, T_2))$ is that either the subspace of T or its orthogonal complement be contained in the union of the two-dimensional subspaces associated with T_1, T_2. The condition is also necessary if* $\dim \mathfrak{X} > 6$ *or if T_1, T_2 intersect.*

Proof. There is no loss in assuming T_i minimal with subspace $\mathfrak{M}_i$ $(i = 1, 2)$. Let T be an involution in $\mathfrak{G}$ which satisfies the given condition and let $T' \varepsilon c(T_1, T_2)$. By Lemma 3. 1, $\mathfrak{M}_1 + \mathfrak{M}_2$ is contained either in the subspace of T' or its orthogonal complement. Thus either the subspace of T or its orthogonal complement is contained either in the subspace of T' or its orthogonal complement. By Lemma 1. 2, this implies that T, T' commute. Therefore $T \varepsilon c(c(T_1, T_2))$ and the sufficiency is proved.

Now assume $\dim \mathfrak{X} > 6$ and let $\mathfrak{N}$ denote any complement of $\mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp$ in $\mathfrak{M}_1{}^\perp$. Since $\mathfrak{M}_2$ is a complement of $\mathfrak{M}_2{}^\perp$ in $\mathfrak{X}$ and $\dim \mathfrak{M}_2 = 2$, it follows that $\dim \mathfrak{N} \leq 2$. We have $\mathfrak{X} = \mathfrak{M}_1 \oplus (\mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp) \oplus \mathfrak{N}$. Hence

$\dim \mathfrak{X} - 2 \leq \dim (\mathfrak{M}_1 \oplus (\mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp))$. Suppose $\mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Then $\dim (\mathfrak{M}_1 \oplus (\mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp)) \leq \dim (\mathfrak{M}_1 + \mathfrak{M}_2) \leq 4$. This implies $\dim \mathfrak{X} \leq 6$ contrary to hypothesis. Therefore $\mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Now take $T \varepsilon c(c(T_1, T_2))$ and let $\mathfrak{M}$ be the subspace of T. Also let y, z be arbitrary elements of $\mathfrak{X}$ such that $y, z \not\varepsilon \mathfrak{M}_1 + \mathfrak{M}_2$ and $z \varepsilon \mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp$. Since $\mathfrak{M}_1 + \mathfrak{M}_2$ is finite-dimensional and $y, z \varepsilon \mathfrak{M}_1 + \mathfrak{M}_2$, there exists $s \varepsilon \mathfrak{X}$ such that $(\mathfrak{M}_1 + \mathfrak{M}_2, s) = (0)$, $(z, s) = 1$ and $(y, s) \neq 0$. Define Z by $xZ = 2(x, s)z - 2(x, z)s$. Then $Z \varepsilon c(T_1, T_2)$. Therefore Z, T commute. By Lemma 1. 2, either $\mathfrak{M}$ or $\mathfrak{M}^\perp$ must contain both s and z. Suppose $s, z \varepsilon \mathfrak{M}$. Then $(\mathfrak{M}^\perp, s) = (\mathfrak{M}^\perp, z) = (0)$. Therefore $y \not\varepsilon \mathfrak{M}^\perp$. Holding z fixed and letting y vary over all vectors not in $\mathfrak{M}_1 + \mathfrak{M}_2$, we see that $\mathfrak{M}^\perp \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Similarly, if $z \varepsilon \mathfrak{M}^\perp$, then we obtain $\mathfrak{M} \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. This proves the necessity when $\dim \mathfrak{X} > 6$. Now the condition $\dim \mathfrak{X} > 6$ was used only to ensure that $\mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. If T_1, T_2 intersect, then $\dim (\mathfrak{M}_1 + \mathfrak{M}_2) = 3$ so that $\mathfrak{M}_1{}^\perp \cap \mathfrak{M}_2{}^\perp$ cannot be contained in $\mathfrak{M}_1 + \mathfrak{M}_2$ even when $\dim \mathfrak{X} = 6$. Therefore the above argument applies and the proof is complete.

LEMMA 3. 3. *Let T_i $(i = 1, 2, 3)$ be non-commutative minimal involutions in $\mathfrak{G}$ which intersect by pairs and let $\mathfrak{M}_i$ be the subspace of T_i. Then always* $\dim (\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 3$ *or* 4 *and is equal to* 3 *if, and only if,* $T_3 \varepsilon c(c(T_1, T_2))$.

Proof. That $\dim (\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 3$ or 4 is obvious. Moreover, if $\dim (\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 3$, then $\mathfrak{M}_3 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$ and $T_3 \varepsilon c(c(T_1, T_2))$ by Lemma 3. 2. Conversely, if $T_3 \varepsilon c(c(T_1, T_2))$, then, again by Lemma 3. 2. $\mathfrak{M}_3 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$ and hence $\dim (\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 3$.

THEOREM 3. 4. *Let T_1, T_2 be non-commutative extremal involutions in $\mathfrak{G}$. Then a necessary and sufficient condition for T_1, T_2 to intersect is that $c(c(S_1, S_2)) = c(c(T_1, T_2))$ for every pair S_1, S_2 of non-commutative extremal involutions in $c(c(T_1, T_2))$.*

Proof. There is obviously no loss in assuming the T_i to be minimal with subspaces $\mathfrak{M}_i$ and written in the form $xT_i = 2(x, u_i)t_i - 2(x, t_i)u_i$, $(t_i, u_i) = 1$ $(i = 1, 2)$. Assume first that T_1, T_2 do not intersect so that $\dim (\mathfrak{M}_1 + \mathfrak{M}_2) = 4$. Define $s_1 = t_1 + t_2 - (t_1, u_2)t_2$. Then s_1, t_2, u_2 are linearly independent and $(s_1, u_2) = 1$. Define $xS_1 = 2(x, u_2)s_1 - 2(x, s_1)u_2$ and take $S_2 = T_2$. Then S_1 is an involution in $\mathfrak{G}$. Moreover, if $\mathfrak{N}_i$ is the subspace of S_i, then $\mathfrak{N}_i \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Therefore $S_i \varepsilon c(c(T_1, T_2))$, by Lemma 3. 2. On the other hand, if $\mathfrak{M}_1 \subseteq \mathfrak{N}_1 + \mathfrak{N}_2$, then $\mathfrak{N}_1 + \mathfrak{N}_2$ contains t_1, u_1, t_2, u_2 and hence coincides with $\mathfrak{M}_1 + \mathfrak{M}_2$. But $\dim (\mathfrak{N}_1 + \mathfrak{N}_2) = 3$ and

$\dim(\mathfrak{M}_1 + \mathfrak{M}_2) = 4$. Therefore $\mathfrak{M}_1 \not\subseteq \mathfrak{N}_1 + \mathfrak{N}_2$. Since S_1, S_2 intersect, this implies, by Lemma 3. 2, that $T_1 \not\varepsilon c(c(S_1, S_2))$ and proves the sufficiency.

Now assume that T_1, T_2 intersect so that $\dim(\mathfrak{M}_1 + \mathfrak{M}_2) = 3$. Let S_1, S_2 be any pair of non-commutative extremal involutions in $c(c(T_1, T_2))$. We can evidently assume that S_i is minimal with subspace $\mathfrak{N}_i$ $(i = 1, 2)$. Since T_1, T_2 intersect, we have, by Lemma 3. 2, that $\mathfrak{N}_1 + \mathfrak{N}_2 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Since S_1, S_2 do not commute, $\dim(\mathfrak{N}_1 + \mathfrak{N}_2) \geq 3$. Therefore $\mathfrak{N}_1 + \mathfrak{N}_2 = \mathfrak{M}_1 + \mathfrak{M}_2$ and, by Lemma 3. 2, it follows that $c(c(S_1, S_2)) = c(c(T_1, T_2))$. This completes the proof.

We are now in a position to use the group isomorphism $G \to \mathfrak{g}(G)$ to construct a one-to-one mapping of the one-dimensional subspaces of $\mathfrak{X}$ onto the one-dimensional subspaces of $\mathfrak{Y}$. Observe first that, if T is an extremal involution in $\mathfrak{G}$ with associated non-isotropic, two-dimensional subspace $\mathfrak{M}$, then $\mathfrak{g}(T)$ is extremal, so determines a non-isotropic, two-dimensional subspace of $\mathfrak{Y}$ which we denote by $\mathfrak{M}\bar{F}$. Evidently $\bar{F}$ is a one-to-one mapping of the non-isotropic two-dimensional subspaces of $\mathfrak{X}$ onto those of $\mathfrak{Y}$. Furthermore, by Theorem 3. 4, $\bar{F}$ preserves the intersection properties of these subspaces.

THEOREM 3. 5. *There exists a one-to-one mapping F of the one-dimensional subspaces of $\mathfrak{X}$ onto the one-dimensional subspaces of $\mathfrak{Y}$ with the properties*:

(i) *If $\mathfrak{M}$ is a non-isotropic two-dimensional subspaces of $\mathfrak{X}$, then $(\mathfrak{D}u)F \subseteq \mathfrak{M}\bar{F}$ if, and only if, $\mathfrak{D}u \subseteq \mathfrak{M}$.*

(ii) *$(\mathfrak{D}u)F \perp (\mathfrak{D}v)F$ if, and only if, $\mathfrak{D}u \perp \mathfrak{D}v$.*

(iii) *$(\mathfrak{D}u_3)F \subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}_2)F$ if, and only if, $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$.*

Proof. Let $\mathfrak{D}u$ be an arbitrary one-dimensional subspace of $\mathfrak{X}$. Choose $u_1 \varepsilon \mathfrak{X}$ such that $(u, u_1) = 1$. Since $\dim(\mathfrak{D}u)^\perp \geq 5$, there exists $u'_1 \varepsilon (\mathfrak{D}u)^\perp$ such that u, u_1, u'_1 are linearly independent. Define $u_2 = u_1 + u'_1$. Then $(u, u_2) = 1$ and u, u_1, u_2 are linearly independent. Let $\mathfrak{M}_i = \mathfrak{D}u + \mathfrak{D}u_i$. Then $\mathfrak{M}_i$ is a non-isotropic, two-dimensional subspace of $\mathfrak{X}$ and $\mathfrak{M}_1 \cap \mathfrak{M}_2 = \mathfrak{D}u$. It follows that $\mathfrak{M}_1\bar{F}$ and $\mathfrak{M}_2\bar{F}$ intersect in a one-dimensional subspace of $\mathfrak{Y}$ which we denote by $(\mathfrak{D}u)F$. The first problem is to show that $(\mathfrak{D}u)F$ is single-valued; that is, that $(\mathfrak{D}u)F$ is independent of the choice of $\mathfrak{M}_1$, $\mathfrak{M}_2$. This result will be established if we prove that a non-isotropic, two-dimensional subspace $\mathfrak{M}$ in $\mathfrak{X}$ contains $\mathfrak{D}u$ if, and only if, $\mathfrak{M}\bar{F}$ contains $(\mathfrak{D}u)F$. This will also establish property (i). We can obviously assume $\mathfrak{M} \neq \mathfrak{M}_i$ $(i = 1, 2)$ and, because of symmetry, it will be sufficient to prove that $\mathfrak{D}u \subseteq \mathfrak{M}$

implies $(\mathfrak{D}u)F \subseteq \mathfrak{M}\bar{F}$. If dim $(\mathfrak{M} + \mathfrak{M}_1 + \mathfrak{M}_2) = 4$, then it follows from Lemma 3. 3 that dim $(\mathfrak{M}\bar{F} + \mathfrak{M}_1\bar{F} + \mathfrak{M}_2\bar{F}) = 4$. Since $\mathfrak{M}\bar{F}$, $\mathfrak{M}_1\bar{F}$, $\mathfrak{M}_2\bar{F}$ intersect by pairs and $\mathfrak{M}_1\bar{F} \cap \mathfrak{M}_2\bar{F} = (\mathfrak{D}u)F$, we obtain $(\mathfrak{D}u)F \subseteq \mathfrak{M}\bar{F}$. Now let dim $(\mathfrak{M} + \mathfrak{M}_1 + \mathfrak{M}_2) = 3$ and choose $r \varepsilon (\mathfrak{D}u)^{\perp}$ such that u, u_1, u_2, r are linearly independent. Define $u_3 = u_1 + r$; then $(u, u_3) = 1$ and u, u_1, u_2, u_3 are linearly independent. If $\mathfrak{M}_3 = \mathfrak{D}u + \mathfrak{D}u_3$, then $\mathfrak{M}_3$ is non-isotropic and dim $(\mathfrak{M} + \mathfrak{M}_2 + \mathfrak{M}_3) = $ dim $(\mathfrak{M} + \mathfrak{M}_1 + \mathfrak{M}_3) = 4$. By the preceding argument, $(\mathfrak{D}u)F \subseteq \mathfrak{M}_3\bar{F}$ and this implies, again by the same argument, that $(\mathfrak{D}u)F \subseteq \mathfrak{M}\bar{F}$.

Next let $\mathfrak{D}u$, $\mathfrak{D}v$ be non-orthogonal one-dimensional subspaces of $\mathfrak{X}$. Then $(u, v) \neq 0$ and the two-dimensional subspace $\mathfrak{M} = \mathfrak{D}u + \mathfrak{D}v$ is non-isotropic. By property (i), we have $\mathfrak{M}\bar{F} = (\mathfrak{D}u)F + (\mathfrak{D}v)F$. Since $\mathfrak{M}\bar{F}$ is non-isotropic, $(\mathfrak{D}u)F$ and $(\mathfrak{D}v)F$ cannot be orthogonal. This result plus symmetry proves (ii).

In order to prove (iii), it will be sufficient to prove that $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$ implies $(\mathfrak{D}u_3)F \subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$. Suppose on the contrary that $(\mathfrak{D}u_3)F \not\subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$ and choose a non-zero vector $v_i \varepsilon (\mathfrak{D}u_i)F$. Evidently v_1, v_2, v_3 are linearly independent so that there exists $v_4 \varepsilon \mathfrak{Y}$ such that $(v_1, v_4) = (v_2, v_4) = 0$ while $(v_3, v_4) = 1$. Let $\mathfrak{D}w_4$ be chosen so that $(\mathfrak{D}u_4)F = \mathcal{E}v_4$. Then, since $\mathcal{E}v_4$ is orthogonal to $(\mathfrak{D}u_1)F$ and $(\mathfrak{D}u_2)F$ but not to $(\mathfrak{D}u_3)F$, it follows, by property (ii), that $\mathfrak{D}u_4$ is orthogonal to $\mathfrak{D}u_1$ and $\mathfrak{D}u_2$ but not to $\mathfrak{D}u_3$. Since this result is inconsistent with $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$, the proof is complete.

4. Representation of the group isomorphism. We prove here that the group isomorphism $G \to \mathfrak{g}(G)$, in both the unitary and symplectic cases, is essentially generated by a linear space isomorphism of $\mathfrak{X}$ onto $\mathfrak{Y}$. The two cases are considered simultaneously.

By a *linear space isomorphism* of $\mathfrak{X}$ onto $\mathfrak{Y}$ we mean a one-to-one semi-linear mapping $x \to x\Phi$ of $\mathfrak{X}$ onto $\mathfrak{Y}$. In other words, $x \to x\Phi$ is additive and, for all $\lambda \varepsilon \mathfrak{D}$ and $x \varepsilon \mathfrak{X}$, $(\lambda x)\Phi = \lambda^{\phi}(x\Phi)$, where $\lambda \to \lambda^{\phi}$ is an isomorphic mapping of the division ring $\mathfrak{D}$ onto $\mathcal{E}$ which is independent of x. The isomorphism Φ is said to *preserve orthongonality* provided $(x\Phi, y\Phi) = 0$ if, and only if, $(x, y) = 0$.

THEOREM 4. 1. *Let $x \to x\Phi$ be a linear space isomorphism of $\mathfrak{X}$ onto $\mathfrak{Y}$ which preserves orthogonality. Then there exists a constant $\alpha \varepsilon \mathcal{E}$ such that $\alpha^* = \alpha$ and*

(i) $\lambda^{\phi *} = \alpha^{-1}\lambda^{*\phi}\alpha$, *for all* $\lambda \varepsilon \mathfrak{D}$,

(ii) $(x\Phi, y\Phi) = (x, y)^{\phi}\alpha$, *for all* $x, y \in \mathfrak{X}$,

(iii) $G \in \mathfrak{U}(\mathfrak{X})$ *implies* $\Phi^{-1}G\Phi \in \mathfrak{U}(\mathfrak{Y})$.

Proof. Let x, u be arbitrary elements of $\mathfrak{X}$ with $u \neq 0$ and choose v such that $(v, u) = 1$. If $x' = x - (x, u)v$, then $(x', u) = 0$. Therefore $(x'\Phi, u\Phi) = 0$, and hence $(x\Phi, u\Phi) = (x, u)^{\phi}(v\Phi, u\Phi)$. Define $\alpha_u = (v\Phi, u\Phi)$ and observe that α_u depends on u but not on x. Now let w be any other non-zero element of $\mathfrak{X}$. By additivity we have $(x, u)^{\phi}(\alpha_u - \alpha_{u+w}) = (x, w)^{\phi}(\alpha_{u+w} - \alpha_w)$, for all x. If u, w are linearly independent, it is immediate that $\alpha_u = \alpha_{u+w} = \alpha_w$ and hence that $\alpha_u = \alpha_w$. If u, w are linearly dependent, choose r linearly independent of u, w. Then $\alpha_u = \alpha_r$ and $\alpha_r = \alpha_w$ so that $\alpha_u = \alpha_w$ in this case as well. It follows that α_u is a constant which we denote by α. Thus $(x\Phi, y\Phi) = (x, y)^{\phi}\alpha$ for all x, y. This proves (ii). If $(v, u) = 1$, then $\alpha = (v\Phi, u\Phi) = \epsilon(u\Phi, v\Phi)^* = \epsilon\alpha^*(u, v)^{\phi *} = \alpha^*(v, u)^{\phi *} = \alpha^*$. That is, $\alpha = \alpha^*$. Moreover, for arbitrary $\lambda \in \mathfrak{D}$, $(v\Phi, (\lambda u)\Phi) = (v, \lambda u)^{\phi}\alpha = (v, u)^{\phi}\lambda^{*\phi}\alpha = \lambda^{*\phi}\alpha$. Also $(v\Phi, (\lambda u)\Phi) = (v\Phi, \lambda^{\phi}u\Phi) = (v\Phi, u\Phi)\lambda^{\phi *} = \alpha\lambda^{\phi *}$. Therefore $\lambda^{\phi *} = \alpha^{-1}\lambda^{*\phi}\alpha$, which is property (i). If $G \in \mathfrak{U}(\mathfrak{X})$, then, for $x, y \in \mathfrak{Y}$,

$$
\begin{aligned}
(x\Phi^{-1}G\Phi, y\Phi^{-1}G\Phi) &= (x\Phi^{-1}G, y\Phi^{-1}G)^{\phi}\alpha \\
&= (x\Phi^{-1}G, y\Phi^{-1})^{\phi}\alpha + (x\Phi^{-1}, y\Phi^{-1}G)^{\phi}\alpha \\
&= (x\Phi^{-1}G\Phi, y) + (x, y\Phi^{-1}G\Phi).
\end{aligned}
$$

Therefore $\Phi^{-1}G\Phi \in \mathfrak{U}(\mathfrak{Y})$. This completes the proof.

LEMMA 4.2. *There exists a linear space isomorphism* Φ *of* $\mathfrak{X}$ *onto* $\mathfrak{Y}$ *which preserves orthogonality and is such that* $(\mathfrak{D}u)\Phi = (\mathfrak{D}u)F$, *where* F *is the mapping given in Theorem 2.6 or 3.5.*

Proof. The existence of a linear space isomorphism $x \to x\Phi$ such that $(\mathfrak{D}u)\Phi = (\mathfrak{D}u)F$ is given, in view of property (iii) in Theorems 2.6 and 3.5, by a known result from projective geometry.[11] That Φ preserves orthogonality follows from property (ii) of Theorems 2.6 and 3.5.

In the next theorem, which is the desired representation theorem, Φ will denote the linear space isomorphism of $\mathfrak{X}$ onto $\mathfrak{Y}$ given by Lemma 4.2. Also $\mathcal{E}_0$ will denote the group, under the circle operation, consisting of all elements μ in the center of $\mathcal{E}$ such that $\mu \circ \mu^* = 0$.

THEOREM 4.3. *The group isomorphism* $G \to \mathfrak{g}(G)$ *has the form*

[11] For an outline of a proof of this result, see, e.g., [7, Lemma 4.4].

$$(4.1) \qquad \mathfrak{g}(G) = \gamma(G)I \circ (\Phi^{-1}G\Phi),$$

where $G \rightarrow \gamma(G)$ *is a homomorphism of* $\mathfrak{G}$ *into* $\mathcal{E}_0$.

Proof. Let T be an arbitrary minimal involution in $\mathfrak{G}$. Then $\Phi^{-1}T\Phi$ is a minimal involution in $\mathfrak{U}(\mathfrak{Y})$ but not necessarily in $\mathfrak{H}$. However the minimal non-isotropic subspace of $\mathfrak{Y}$ associated with $\Phi^{-1}T\Phi$ is the same as that associated with $\mathfrak{g}(T)$. Therefore $\mathfrak{g}(T) = \rho I \circ (\Phi^{-1}T\Phi)$, where $\rho = 0$ or 2. Hence $\mathfrak{g}$ has the form (4.1) for minimal involutions. Now define the mapping

$$(4.2) \qquad G \rightarrow G^{\sigma} = \Phi\mathfrak{g}(G)\Phi^{-1},$$

which is clearly an isomorphism of $\mathfrak{G}$ onto another subgroup of $\mathfrak{U}(\mathfrak{X})$. Observe that, if T is a minimal involution in $\mathfrak{G}$, then $T^{\sigma} = \rho I \circ T$, where $\rho = 0$ or 2.

Consider an arbitrary $G \varepsilon \mathfrak{G}$ and an arbitrary minimal involution $T \varepsilon \mathfrak{G}$. Then $G \circ T \circ G^{\circ} = (I - G)T(I - G)^{-1}$ is also a minimal involution. Therefore $(G \circ T \circ G^{\circ})^{\sigma} = \rho_1 I \circ (G \circ T \circ G^{\circ})$, $\rho_1 = 0$ or 2. But also $(G \circ T \circ G^{\circ})^{\sigma} = G^{\sigma} \circ T^{\sigma} \circ (G^{\sigma})^{\circ}$ and $T^{\sigma} = \rho_2 I \circ T$, $\rho_2 = 0$ or 2. Hence $G^{\sigma} \circ T \circ (G^{\sigma})^{\circ} = \rho I \circ (G \circ T \circ G^{\circ})$, $\rho = \rho_1 \circ \rho_2 = 0$ or 2. This last result can be written in the form

$$\rho I = (I - G^{\sigma})T(I - G^{\sigma})^{-1} + (I - G)T(I - G)^{-1}(\rho - 1).$$

Each term on the right is at most two-dimensional and consequently the right hand side is at most four-dimensional. Since $\dim \mathfrak{X} \geq 6$, it follows that $\rho = 0$. In other words,

$$(4.3) \qquad (I - G^{\sigma})T(I - G^{\sigma})^{-1} = (I - G)T(I - G)^{-1}.$$

Suppose now that there exists a $z \varepsilon \mathfrak{X}$ such that $v = z(I - G^{\sigma})$ and $u = z(I - G)$ are linearly independent. We prove that this leads to a contradiction. In the unitary case choose, by Lemma 2.2, a non-isotropic vector w such that $(u, w) = 1$ and $(v, w) = 0$. Define $xT = 2(x, w)(w, w)^{-1}w$. Then $uT \neq 0$ and $vT = 0$. With this choice of T, apply (4.3) to z to obtain $uT(I - G)^{-1} = 0$. Since this implies $uT = 0$, we have a contradiction in the unitary case. In the symplectic case, choose w linearly independent of u, v such that $(u, w) = (v, w) = 0$ and then choose w' such that $(u, w') = (w, w') = 1$ and $(v, w') = 0$. Define $xT = 2(x, w')w - 2(x, w)w'$. Then $uT = 2w$ and $vT = 0$. With this choice of T, apply (4.3) to z to obtain $2w(I - G)^{-1} = 0$. This implies $w = 0$, contrary to the choice of w. We

have thus proved that $x(I-G^{\sigma})$ and $x(I-G)$ are linearly dependent for all x. A similar proof shows that $x(I-G^{\sigma})$ and $x(I-G)$ also vanish simultaneously. Therefore, for every $x \varepsilon \mathfrak{X}$, there exists $\lambda_x \varepsilon \mathfrak{D}$ such that $x(I-G^{\sigma})=\lambda_x x(I-G)$ or

$$x(I-G^{\sigma})(I-G)^{-1}=\lambda_x x. \tag{4.4}$$

Since the left hand side of (4.4) is additive in x, we obtain for all x, y,

$$(\lambda_{x+y}-\lambda_x)x+(\lambda_{x+y}-\lambda_y)y=0. \tag{4.5}$$

Since $\dim \mathfrak{X}>1$, (4.5) implies that λ_x is equal to a constant λ independent of x. Hence $I-G^{\sigma}=\lambda(I-G)$ and this can be written in the form $G^{\sigma}=(1-\lambda)I\circ G$. Returning to $\mathfrak{g}(G)$, we have

$$\mathfrak{g}(G)=\gamma(G)I\circ(\Phi^{-1}G\Phi), \tag{4.6}$$

where $\gamma(G)=(1-\lambda)^{\phi}$. It remains to prove that $G\rightarrow\gamma(G)$ is a homomorphism of $\mathfrak{G}$ into $\mathcal{E}_0$. It is obvious from the linearity of $\mathfrak{g}(G)$ that $\gamma(G)$ is in the center of $\mathcal{E}$. Moreover, for all $x, y \varepsilon \mathfrak{Y}$,

$$(x\mathfrak{g}(G), y\mathfrak{g}(G))=(x\mathfrak{g}(G), y)+(x, y\mathfrak{g}(G)). \tag{4.7}$$

Substituting (4.6) in (4.7) with $\gamma=\gamma(G)$, we obtain

$$\begin{aligned}\gamma(x,y)\gamma^*&+\gamma(x,y\Phi^{-1}G\Phi)(1-\gamma^*)+(1-\gamma)(x\Phi^{-1}G\Phi,y)\gamma^*\\&+(1-\gamma)(x\Phi^{-1}G\Phi,y\Phi^{-1}G\Phi)(1-\gamma^*)\\&=\gamma(x,y)+(1-\gamma)(x\Phi^{-1}G\Phi,y)+(x,y)\gamma^*+(x,y\Phi^{-1}G\Phi)(1-\gamma^*).\end{aligned} \tag{4.8}$$

Using the fact that $\Phi^{-1}G\Phi \varepsilon \mathfrak{U}(\mathfrak{Y})$ and that γ is in the center of $\mathcal{E}$ and then collecting terms in (4.8), we finally obtain $\gamma\circ\gamma^*=0$. Hence $\gamma \varepsilon \mathcal{E}_0$. It is obvious that $G\rightarrow\gamma(G)$ is a homomorphism of $\mathfrak{G}$ into $\mathcal{E}_0$; therefore the proof is complete.

COROLLARY. *If $\mathfrak{G}$ coincides with its commutator subgroup, then $\gamma(G)\equiv 0$. In particular, if $\mathfrak{X}$ is symplectic and finite-dimensional, then $\gamma(G)\equiv 0$* [4, p. 12].

Proof. Since $\mathcal{E}_0$ is commutative, γ maps commutators of $\mathfrak{G}$ into 0.

COROLLARY. *If $\mathcal{E}$ is a field with the identity mapping as involution, then $\gamma(G)=0$ or 2 for all $\mathfrak{G}$.*

Proof. In this case we have $\gamma(G)\circ\gamma(G)=0$ which implies $\gamma(G)=0$ or 2.

We close with a remark on the unitary case when the vector spaces have index zero; that is, every vector is non-isotropic. This case is of course included in the above treatment; however the discussion becomes considerably simpler. Moreover it is possible to weaken the dimension condition from six to three and drop the condition which excludes the field F_3.

YALE UNIVERSITY.

BIBLIOGRAPHY.

[1] J. Dieudonné, "On the automorphisms of the classical groups," *Memoirs of the American Mathematical Society*, No. 2 (1950).

[2] ———, "Sur les automorphismes des groupes classiques," *Comptes Rendus de l'Académie des Sciences (Paris)*, vol. 225 (1947), pp. 914-915.

[3] ———, "Sur les automorphismes du group unitaire," *ibid.*, vol. 225 (1947), pp. 975-976.

[4] ———, "Sur les groupes classiques," *Actualités Scientifiques et Industrielles*, No. 1040 (1948).

[5] L. K. Hua, "On the automorphisms of the symplectic group over any field," *Annals of Mathematics*, vol. 49 (1948), pp. 739-759.

[6] N. Jacobson, "On the theory of primitive rings," *ibid.*, vol. 48 (1947), pp. 8-21.

[7] C. E. Rickart, "Isomorphic groups of linear transformations," *American Journal of Mathematics*, vol. 72 (1950), pp. 451-464.

ON THE IRREDUCIBILITY OF POLYNOMIALS WITH LARGE THIRD COEFFICIENT. II.*

By Alfred Brauer.

This paper is a continuation of my paper "On the irreducibility of polynomials with large third coefficient," this Journal, vol. 70 (1948), pp. 423-432. The enumeration of the theorems and equations will be continued here. Combining the main results of the first paper and new results we will prove the following theorem.

Theorem 5. *Let $f(x) = x^n + a_1x^{n-1} + \cdots + a_n$ $(a_n \neq 0)$ be a polynomial with integral rational coefficients. Let m and m^* be the minimum of the partial sums of the series $0 + a_3 + a_4 + \cdots + a_n$ and $0 - a_3 + a_4 - + \cdots + (-1)^n a_n$, respectively.*

Assume that one of the following conditions is satisfied:

(α) $a_2 \geqq a_1^2/4 + |m^*|$ for $n > 2$ and $a_1 > 0$,

(β) $a_2 \geqq a_1^2/4 + |m|$ for $n > 2$ and $a_1 < 0$,

(γ) $|a_1| \leqq 3$,

(δ) The inequality

$$|a_\nu| \geqq \tfrac{1}{4}\binom{n-2}{\nu-2}\{|a_1| + n - 2\}^2 + \binom{n-2}{\nu-1}\{|a_1| + n - 2\} + \binom{n-2}{\nu}$$

holds for at least one value of ν $(\nu = 2, 3, \cdots, n)$.

We set $t = |1 + a_4| + |a_1 + a_3| + |a_5| + |a_6| + \cdots + |a_n|$. If

$$a_2 > t, \tag{34}$$

then $f(x)$ is irreducible in the field of rational numbers.

Proof. The cases (α) and (β) follow from the proof of Theorem 1a. It follows from (34) as in the proof of Theorem 1 that $f(x)$ has $n - 2$ roots in the interior and 2 roots in the exterior of the unit circle, and for the irreducibility of $f(x)$ it is sufficient to prove that the two roots in the exterior of the unit circle are imaginary roots.

* Received July 16, 1950.

Let us assume that these two roots are real. We denote them by x_1 and x_2, and the other roots by $x_3, x_4, \cdots, x_n$.

Assume now that (γ) holds. For $a_1 = 0$ the statement follows from Theorem 2, but it will be obtained here again. It is sufficient to prove that $f(x)$ has no real root greater than or equal to 1. It follows then that $f(-x)$ has no such root either since its coefficients satisfy (γ) and (34). Therefore $f(x)$ has no real root less than or equal to -1, and x_1 and x_2 are imaginary roots.

We have to consider some cases.

1). $a_1 \geqq 0$. It follows from (34) that

$$a_2 > |a_4| - 1 + |a_3| - |a_1| + |a_5| + |a_6| + \cdots + |a_n|,$$

hence for $x \geqq 1$

$$a_2x^{n-2} > -x^{n-2} - a_1x^{n-2} + (|a_3| + |a_4| + \cdots + |a_n|)x^{n-2},$$

$$x^n + a_1x^{n-1} + a_2x^{n-2} > (|a_3| + |a_4| + \cdots + |a_n|)x^{n-2},$$

$$f(x) \geqq x^n + a_1x^{n-1} + a_2x^{n-2} - (|a_3| + |a_4| + \cdots + |a_n|)x^{n-2} > 0.$$

2). $-1 \geqq a_1 \geqq -3$ and $a_3 \leqq 0$. It follows from (34) that

$$a_2 - 1 \geqq |a_1| + |a_3| + |a_4| - 1 + |a_5| + \cdots + |a_n|,$$

hence for $x \geqq 1$

$$(35) \qquad a_2x^{n-2} \geqq (|a_1| + |a_3| + |a_4| + \cdots + |a_n|)x^{n-2}.$$

On the other hand

$$(36) \qquad x^n - |a_1|\, x^{n-1} + |a_1|\, x^{n-2} = x^{n-2}(x^2 - |a_1|\, x + |a_1|) > 0,$$

since $a_1 = -1, -2$ or -3. By adding (35) and (36) we obtain

$$x^n + a_1x^{n-1} + a_2x^{n-2} > (|a_3| + |a_4| + \cdots |a_n|)x^{n-2},$$

hence $f(x) > 0$ for $x \geqq 1$.

3). $-1 \geqq a_1 \geqq -3$ and $a_3 > 0$.

We have for $x \geqq 1$

$$(37) \quad x^n - 3x^{n-1} + x^{n-2} + 3x^{n-3} - x^{n-4} = x^{n-4}\{x^2(x-2)^2 + (x-1)^3\} > 0$$

and

$$(38) \qquad x^{n-1} - x^{n-3} \geqq 0.$$

It follows from (37) and (38) that

(39) $$x^n + a_1 x^{n-1} + x^{n-2} + |a_1| x^{n-3} - x^{n-4} > 0$$

for $-1 \geqq a_1 \geqq -3$. Moreover by (34)

(40) $$a_2 - 1 \geqq |a_1| - |a_3| + |1 + a_4| + |a_5| + |a_6| + \cdots + |a_n|,$$

(41) $$(a_2 - 1)x^{n-2} \geqq (a_2 - 1)x^{n-3} \geqq (|a_1| - |a_3|)x^{n-3} + |1 + a_4| x^{n-4} + (|a_5| + |a_6| + \cdots + |a_n|)x^{n-4}.$$

Adding (41) and (39) we obtain

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} + a_3 x^{n-3} > (|a_4| + |a_5| + \cdots + |a_n|)x^{n-4}.$$

Hence $f(x) > 0$ for $x \geqq 1$, and the theorem is proved if (γ) holds.

If $a_1 = 4$, then the polynomial may have real roots in the exterior of the unit circle. For instance, $x^5 - 4x^4 + 2x^3 + 4x^2 - x - 1$ is negative for $x = 2$, although its coefficients satisfy (34).

Now we assume that (δ) holds. It follows from (34) that

$$\begin{aligned} f(1) &= 1 + a_1 + a_2 + \cdots + a_n \\ &\geqq a_2 - |1 + a_4| - |a_1 + a_3| - |a_5| - |a_6| - \cdots - |a_n| \\ &= a_2 - t > 0. \end{aligned}$$

Therefore either no root or two roots are greater than 1, hence sign $x_1 =$ sign x_2, hence

(42) $$|x_1 x_2| \leqq |x_1 + x_2|^2/4.$$

Let us denote the κ-th elementary symmetric function of $x_3, x_4, \cdots, x_n$ by C_κ. Since $|x_\rho| < 1$ for $\rho = 3, 4, \cdots, n$, we have

(43) $$|C_\kappa| < \binom{n-2}{\kappa} \qquad (\kappa = 1, 2, \cdots, n-2).$$

On the other hand $x_1 + x_2 = -a_1 - C_1$, hence, by (43) and 42), respectively,

(44) $$|x_1 + x_2| < |a_1| + n - 2,$$ (45) $$|x_1 x_2| < (|a_1| + n - 2)^2/4.$$

Moreover $(-1)^\nu a_\nu = x_1 x_2 C_{\nu-2} + (x_1 + x_2) C_{\nu-1} + C_\nu$ for $\nu = 2, 3, \cdots, n$ and by (43), (44), and (45)

$$|a_\nu| < \tfrac{1}{4}\binom{n-2}{\nu-2}\{|a_1| + n - 2\}^2 + \binom{n-2}{\nu-1}\{|a_1| + n - 2\} + \binom{n-2}{\nu}.$$

Since for a certain ν the inequality (δ) holds, a contradiction is obtained.

Therefore x_1 and x_2 are imaginary roots. This completes the proof of the theorem.

Our result can be formulated also as follows:

THEOREM 5a. *If a polynomial with integral rational coefficients satisfies the assumptions of* Theorem 5, *then it has a pair of complex roots in the exterior of the unit circle while all the other roots lie in the interior of the unit circle.*

A similar theorem can be obtained for polynomials with arbitrary real coefficients.

It was shown in the first paper that (34) gives the best possible bound for a_2 if $f(x) = x^5 - 4x^4 + Kx^3 + 4x^2 - 2x - 2$. For these polynomials the condition (β) holds if $K > t$.

Moreover, it was stated there that (β) holds for the polynomials

$$f(x) = x^n - a_1x^{n-1} + a_2x^{n-2} - a_3x^{n-3} - \cdots - a_n \tag{46}$$

where all the a_ν are non-negative, a_4 and a_n positive and

$$|a_1| \leqq 2t^{1/2}. \tag{47}$$

But this is not correct since $m \neq 0$. However, it follows now that (34) gives the best bound for a_2 for the polynomials (46) if instead of (47) either (γ) or (δ) holds. In these cases the polynomials are irreducible for $a_2 > t$ while they are reducible for $a_2 = t$ since $f(1) = a_2 - t = 0$.

In the formulation of Theorem III in the first paper a factor s^2 is erroneously omitted. It should read "or if $a_2 \geqq (7/2)^{2n-2}s^2$."

UNIVERSITY OF NORTH CAROLINA.

SUPPLEMENT TO A PAPER OF G. DE B. ROBINSON.*

By R. M. THRALL and G. de B. ROBINSON.

This note is a supplement to a paper of one of the authors.[1] Its purpose is to correct a formula in that paper, and to derive a new formula which depends on this corrected form.

1. The revisions. The portion of SG_3, § 7, which appears on p. 291 should read as follows:

But we can go deeper still. Let S_b^* be the symmetric group of degree b and let W^* be an element of S_b^* with cycle structure b_1 1-cycles, b_2 2-cycles, $\cdots$, b_k k-cycles. Then let W^{*+} be an element of S_n $(n = a + bq)$ with a 1-cycles, b_1 q-cycles, b_2 $2q$-cycles, $\cdots$. b_k kq-cycles, and let V be any permutation on the fixed letters of W^{*+}. We prove the following generalization of Theorem 6. 5.

7. 1r $$\chi_\alpha(VW^{*+}) = \sigma_\alpha \cdot (\chi(W^*) \text{ in } [\alpha]_q^*) \cdot \chi_{\alpha^0}(V).$$

Now we can write

7. 2r $$\begin{aligned}\chi_\alpha(VW^{*+}) &= \Sigma\chi_\gamma(V) \cdot \chi_\alpha{}^\gamma(P_b \cdot W) \\ &= \Sigma\chi_\gamma(V) \cdot \chi_{(b_1q)}\chi_{(b_2q)} \cdots \chi_{(b_kq)}.\end{aligned}$$

As before, we are only interested in those terms which arise from a sequence of skew hook representations (b_iq), so that we can take $[\gamma] = [\alpha^0]$. Consider one such term on the right hand side of 7. 2r and apply 5. 4 to its component hooks. Here $\sigma = \sigma_\alpha$ is the product of the parities of the hooks of length q, σ' is the product of the parities of the b_iq-hooks, and σ^* is the product of the parities of the corresponding b-hooks in $[\alpha]_q^*$. Io e. multiplying the k equations 5. 4 we obtain

7. 3 $$\sigma' = \sigma\sigma^*.$$

It follows from 5. 2 that σ is the same for all such terms of 7. 2r, so we can sum 7. 3, applying the Murnaghan-Nakayama recursion formula iterated k-times to yield the equation

* Received May 1, 1950.

[1] G. de B. Robinson, "On the representations of the symmetric group" (third paper), *American Journal of Mathematics*, vol. 70 (1948), pp. 277-294. We shall refer to this paper as SG_3.

7.4r $$\chi_a^{a^0}(W^{*+}) = \sigma_a \cdot (\chi(W^*) \text{ in } [\alpha]_q^*),$$

which proves 7.1r.

As before, if $[\alpha']$ is a second diagram for n with q-core containing $a' > a$ nodes then we conclude that

7.5r $$\chi_{a'}(VW^{*+}) = 0.$$

These changes also apply to the paragraph of SG_3 beginning on the bottom of p. 278 and in particular to formula (7) p. 279.

2. The new formula. We are interested in the frequency τ_a of the identity representation of S_b^* in $[\alpha]_q^*$. On the one hand this frequency (SG_3, p. 284, last paragraph) is one or zero according as each disjoint constituent of $[\alpha]_q^*$ does or does not consist of a single row. On the other hand we can get this frequency from the orthogonality relations for characters of S_b^*. Thus we have

$$(\Sigma\chi(W^*) \text{ in } [\alpha]_q^*)/b\,! = \tau_a,$$

where the sum is over all W^* in S_b^*. Now sum 7.2r over W^*, substitute the above formula and we get

(7.9) $$\Sigma\chi_a(VW^{*+})/b\,! = \tau_a\sigma_a\chi_a^0(V).$$

Similarly, if $[\alpha']$ has q-core with more than a nodes we have on summing (7.5r) over W^* that

(7.10) $$\Sigma\chi_{a'}(VW^{*+}) = 0.$$

The case $n = bq$ is of especial interest. Then we shall prove that for every diagram $[\alpha]$ with n nodes

(7.11) $$\sum_{W^*\epsilon S_b^*} \chi_a(W^{*+})/b\,! = \theta_a,$$

where θ_a is zero if the q-core of α is not zero or if $[\alpha]$ has more than q-rows, and $\theta_a = \sigma_a$ if the q-core of $[\alpha]$ is zero and $[\alpha]$ has q or less rows.

Proof. When $n = bq$ we have $a = 0$ and hence (7.10) applies for all $[\alpha]$ whose q-core is not zero. For the $[\alpha]$ with q-core zero we apply (7.9) and the formula (7.11) will be established as soon as we can show that for $[\alpha]$ with zero q-core τ_a is 0 or 1 according as $[\alpha]$ has $k > q$ or $k \leqq q$ rows.

We first state a criterion for a diagram of $n = bq$ nodes to have zero core. Let $[\alpha]$ correspond to the partition $\alpha_1 + \cdots + \alpha_k + \alpha_{k+1} + \cdots + \alpha_m = n$, where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_k > 0 = \alpha_{k+1} = \cdots = \alpha_m$ and $m \geq k$, and let $l_j = \alpha_j + m - j$, $j = 1, \cdots, m$.

7.12. *Let m be a multiple of q. Then the diagram $[\alpha]$ has zero q-core if and only if exactly m/q of the l_j fall into each residue class modulo q.*

This criterion is implicit in the Murnaghan-Nakayama recursion formula [2] as well as in SG_3 and papers of Todd [3] and Staal.[4] We give here a direct proof.

Suppose that $[\alpha']$ is the diagram obtained by removing from $[\alpha]$ the rim of a q-hook which begins in the i-th row and ends in the j-th row. Then for

$$\begin{array}{lll} h < i & \alpha'_h = \alpha_h & l'_h = l_h \\ i \leq h < j & \alpha'_h = \alpha_{h+1} - 1 & l'_h = l_{h+1} \\ h = j & \alpha'_j = \alpha_i - q + j - i & l'_j = l_i - q \\ h > j & \alpha'_h = \alpha_h & l'_h = l_h. \end{array}$$

Thus we see that for any fixed m the congruence class mod q for the l_j are the same as for the l'_j. If s removals of q-hooks from $[\alpha]$ give $[\alpha^{(s)}]$ we see that the congruence classes mod q for the l_j are the same as for the $l_j^{(s)}$. If, in particular, $n = sq$ then $[\alpha^{(s)}]$ is the zero diagram, $l_j^{(s)} = m - j$, and hence the $l_j^{(s)}$ fall into q congruences classes each containing m/q elements. This establishes the necessity of the criterion.

For any m there will be a q-hook beginning at the end of the i-th row if and only if $l_i \geq q$ and $l_i - q$ is not equal to any l_h for $h > i$. For, if $l_i - q$ is non negative and different from the other l_h we can find j such that $l_j > l_i - q > l_{j+1}$. The corresponding α_j must be positive (i. e. $j \leqq k$) since $2 \leq l_j - l_{j+1} = \alpha_j - \alpha_{j+1} + 1$ or $\alpha_j \geq \alpha_{j+1} + 1$. Moreover, the q-th node down along the rim starting from the end of the i-th row will lie in the j-th row and beyond the node above the end of the $(j+1)$-th row, which shows that a q-hook can be removed. If $l_i - q$ is negative or equal to some l_j then either $j \leq k$ and the q-th node down from the end of the i-th row falls in the $(j-1)$-th row directly above the last node of the j-th row or $j > k$ and there are less than q-nodes available from which to form a hook.

Under the hypothesis of the criterion we have $l_{k-1} = m - k - 1$ and $l_k = \alpha_k + m - k$. There must be at least one $h < k$ for which $l_h \equiv m - k$. Let i be the largest such index. Then $l_i = \alpha_i + (k - i) + (m - k)$ so that $l_i - q \geq m - k$, i. e. $l_i - q \neq l_h$ for $h > k$. Because of the maximal nature

[2] F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 59 (1937), pp. 437-488.

T. Nakayama, "On some modular properties of irreducible representations of the symmetric group, Part I," *Japanese Journal of Mathematics*, vol. 17 (1940), pp. 165-184.

[3] J. A. Todd, "A note on the algebra of S-functions," *Proceedings of the Cambridge Philosophical Society*, vol. 45 (1949), pp. 328-334.

[4] R. A. Staal, "Star diagrams and the symmetric group," *Canadian Journal of Mathematics*, vol. 2 (1950), pp. 79-92.

of i we cannot have $l_i - q = l_h$ for $i < h \leqq k$. Hence, there is a q-hook beginning at the end of the i-th row.

Now, that we have shown that the congruence conditions of the criterion guarantee the existence of at least one q-hook it follows from the congruence of the l_j and the l_j' that $[\alpha']$ will also have at least one hook and thus finally leads to the conclusion that $[\alpha]$ has zero q-core, i. e. shows that the criterion affords a sufficient condition for a zero core.

We now return to the determination of τ_α. Let $[\alpha]$ have q-core zero and choose s so that $m = sq \geq k > (s-1)q$. Then none of the non-negative integers l_h for $h > k$ will be as large as $q - 1$, hence exactly s of the l_h with $h \leq k$ will be congruent to $q - 1 \bmod q$. Let these be $l_{i_1} < l_{i_2} < \cdots < l_{i_s}$. We have seen above that there is a q-hook Q_1 beginning at the end of the i-th row and since the l_{i_j} are all congruent mod q, there will be for each j a b_jq-hook Q_j beginning at the end of the i_j-th row and ending with the last node of Q_1.

Now it follows from [5] Theorem 4. 7 of SG_3 that whenever $s > 1$ some disjoint constituent of $[\alpha]_q{}^*$ will have more than one row. On the other hand if $k \leq q$ (i. e., if $s = 1$) no two l_h for $h \leq k$ can be congruent and hence no disjoint constituent of $[\alpha]_q{}^*$ can contain more than one row. This completes the proof of Theorem 7. 11.

The arguments used in the proof of Theorem 7. 12 serve also to give a short proof of the uniqueness of the q-core (first proved by Nakayama [6]). Rephrasing the criterion that a diagram have a q-hook we get the following theorem:

7. 13. *A diagram is a q-core if and only if each class of congruent l_j's (formed for any m) contains all smaller non-negative integers congruent to the largest one in the class.*

Consider two q-cores $[\alpha]$ and $[\alpha']$ and select any $m \geqq \max(k, k')$. Then if each congruence class mod q for the l_j has the same order as the corresponding class for the l_j', it follows from 7. 13 that $[\alpha] = [\alpha']$. But since the congruence classes mod q are the same for a diagram $[\beta]$ as for any diagram obtained from $[\beta]$ by removing q-hooks we conclude that $[\beta]$ can have only one q-core.

University of Michigan
and
University of Toronto.

[5] See also R. A. Staal, *loc. cit.*, Theorem C.

[6] T. Nakayama, *loc. cit.*

ok

on
ce
lly
on

ero
ve
th
l_i.
he
a
ast

ne
er
nd
is

a
).
ng

,'s
to

en
e-
ce
ny
an